

# MASTER'S THESIS

**DPQM, een procesbenadering van datakwaliteit**

**Om problemen met datakwaliteit inzichtelijk te maken, te analyseren en op te lossen**

Philipsen, P.R.A. (Peter)

**Award date:**  
2019

[Link to publication](#)

## **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

## **Take down policy**

If you believe that this document breaches copyright please contact us at:

[pure-support@ou.nl](mailto:pure-support@ou.nl)

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 05. May. 2023

**Open Universiteit**  
[www.ou.nl](http://www.ou.nl)



# DPQM, een procesbenadering van datakwaliteit

Om problemen met datakwaliteit inzichtelijk te maken, te analyseren en op te lossen

## DPQM, a process approach of data quality

To provide insight, analyze and solve problems with data quality



|                     |  |
|---------------------|--|
| Opleiding:          | Open Universiteit, faculteit Management, Science & Technology<br>Masteropleiding Business Process Management & IT          |
| Programme:          | Open University of the Netherlands, faculty of Management, Science & Technology<br>Master Business Process Management & IT |
| Cursus:             | IM0602 Voorbereiden Afstuderen BPMIT<br>IM9806 Afstudeertraject Business Process Management and IT                         |
| Student:            | P.R.A. Philipsen   |
| Identiteitsnummer:  |  |
| Datum:              | 28-04-2019   |
| Afstudeerbegeleider | Prof. dr. ir. R.W. Helms   |
| Meelezer            | Dr. A. D. Counotte-Potman  |
| Versie nummer:      | 2.0  |
| Status:             | Definitief   |



## Abstract

Zowel in de praktijk als in de literatuur wordt het belang van data als key asset in organisaties en de kwaliteit van data onderkend. Datakwaliteit is de mate waarin data bruikbaar is voor het gegeven doel in een gegeven context. Er zijn verschillende situaties van veel voorkomende problemen met datakwaliteit waardoor gegevens uiteindelijk niet gebruikt worden. De belangrijkste strategieën om issues met betrekking tot datakwaliteit op te lossen zijn datagedreven strategieën en procesgedreven strategieën. Vanuit langetermijnperspectief geniet de laatste de voorkeur.

Het doel van dit onderzoek is om organisaties handvatten te geven om problemen met datakwaliteit in businessstermen begrijpelijk te maken, te analyseren en op te kunnen lossen. De vraag van deze afstudeeropdracht is te onderzoeken in hoeverre een datakwaliteitsgeoriënteerde manier van procesmodelleren daarbij kan helpen. Daartoe is DQPM ontwikkeld, een methode met zeven stappen die problemen in kaart brengt, visualiseert in procesmodellen en oplossingen aangeeft. De methode is beproefd met een case. De methode voldoet aan de criteria die verzameld zijn in de literatuur. Er zijn nog wel een aantal punten voor vervolgonderzoek om de methode te verbeteren en de externe validiteit te verhogen.

## Sleutelbegrippen

Datakwaliteit, Business Process Management, BPMN, Requirements, Kwaliteitsdimensies

## Samenvatting

Zowel in de praktijk als in de literatuur wordt het belang van data als key asset in organisaties en de kwaliteit van data onderkend. Datakwaliteit is de mate waarin data bruikbaar is voor het gegeven doel in een gegeven context. Door middel van dimensies kan de kwaliteit van data beoordeeld worden. Dimensies geven metrieke om kwaliteit te kunnen meten. Voorbeelden van kwaliteitsdimensies zijn juistheid, geloofwaardigheid, volledigheid, toegankelijkheid en begrijpelijkheid.

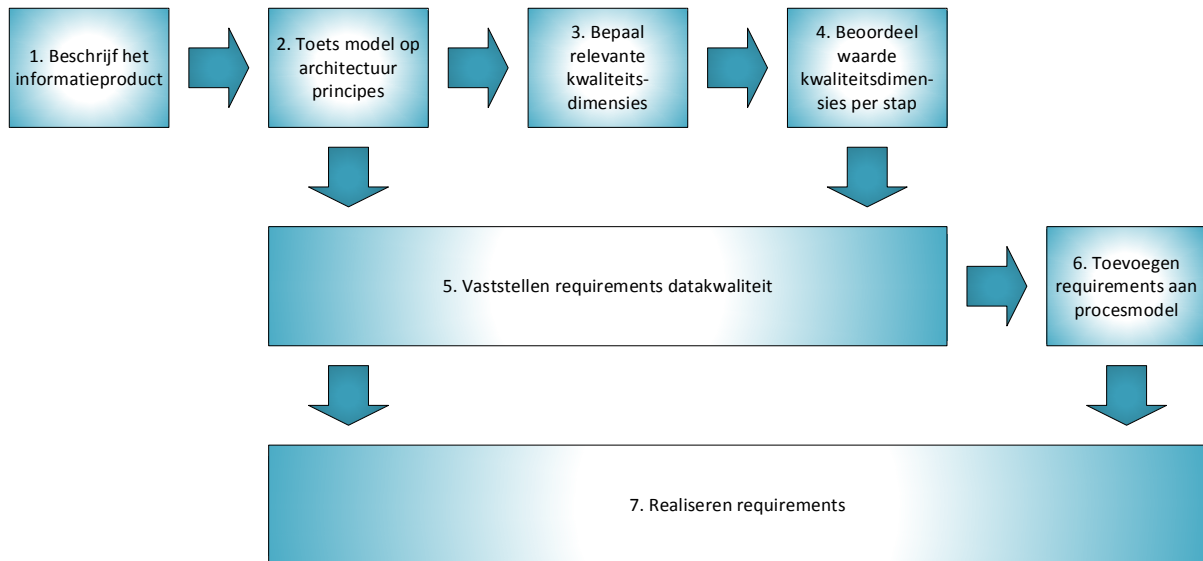
Er zijn verschillende situaties van veel voorkomende problemen met datakwaliteit. Door meerdere bronnen voor dezelfde gegevens wordt de geloofwaardigheid van data nadelig beïnvloed. Dat zorgt ervoor dat de gegevens niet gebruikt worden, waardoor de toegevoegde waarde ervan afneemt. Dit leidt tot het kernprobleem; geen gebruik van gegevens. Voor operationele problemen in dataproductie, veranderende behoeften van afnemers van data en gedistribueerde gegevensverwerking geldt dat er vergelijkbare patronen bekend zijn die er uiteindelijk voor zorgen dat gegevens niet gebruikt worden. De oorzaken van problemen met datakwaliteit zijn divers. Dimensies kunnen elkaar bijvoorbeeld beïnvloeden. Streven naar tijdigheid kan gevolgen hebben voor de volledigheid of juistheid. Daarnaast spelen factoren in het informatieverwerkende proces een rol. Verder zijn datakwaliteitsproblemen onoverkomelijk als systemen nauw verbonden zijn en de componenten van systemen complexe interactiepatronen kennen.

De impact van problemen door datakwaliteit op bedrijfsprocessen kan groot zijn. Bijna alle methoden om datakwaliteit te verbeteren bevatten drie fases. Een beschrijvende fase, een assessmentfase waarin de kwaliteitsdimensies beoordeeld worden, en een verbeterfase waarin strategieën en technieken ingezet worden om een hogere kwaliteit van data te bereiken. De belangrijkste strategieën om issues met betrekking tot datakwaliteit op te lossen zijn datagedreven strategieën en procesgedreven strategieën. De procesgerichte aanpak kent twee hoofdbenaderingen: beheersen van het proces of herontwerp van het proces. Beide benaderingen zorgen ervoor dat problemen voorkomen worden en hebben vanuit langetermijnperspectief de voorkeur boven de datagedreven strategie.

Het doel van dit onderzoek is om organisaties handvatten te geven om problemen met datakwaliteit in business termen begrijpelijk te maken, te analyseren en op te kunnen lossen. De vraag van deze afstudeeropdracht is te onderzoeken in hoeverre een datakwaliteitsgeoriënteerde manier van procesmodelleren daarbij kan helpen. Er zijn meerdere methoden ontwikkeld die dat beogen, maar die voldoen in beperkte mate aan de criteria die stakeholders stellen. Die criteria zijn onder te verdelen in drie rubrieken:

- Uitgangspunten, modellen moeten aansluiten bij standaard Business Process Management (BPM)-methoden en moeten vertrouwd zijn voor stakeholders.
- Algemene eisen, ondermeer dat modellen op een systematische manier processen moeten laten zien die nodig zijn om een informatieproduct voort te brengen, dat deze bruikbaar is als input voor een systeemontwikkelingstraject en dat de grenzen van betrokken organisaties en systemen zichtbaar moeten zijn.
- Eisen gericht op datakwaliteit, waaronder dat kwaliteitscontroles zichtbaar moeten zijn, kritieke fasen die kwaliteit beïnvloeden inzichtelijk zijn en dat het model ondersteunend moet zijn om kwaliteitsmaatregelen te implementeren.

Met deze criteria is DQPM ontwikkeld, Data Quality Process Modelling, een methode bestaand uit zeven stappen. Figuur 1 laat de samenhang zien tussen de stappen.



Figuur 1 Overzicht methode DQPM

De eerste stap beschrijft het informatieproduct, de eisen die afnemers er aan stellen en de plek in het proces waar het product gecreëerd wordt. Dit gebeurt op het niveau van werkproces. In de tweede stap wordt gekeken naar het volledige bedrijfsproces waar het werkproces onderdeel van is en wordt beoordeeld of het model voldoet aan belangrijke architectuurprincipes zoals 'quality at source', het vaststellen van kwaliteit van gegevens aan het begin van het proces. Afhankelijk van de uitkomst levert dit requirements op voor aanpassing van het proces. De volgende stap wordt weer uitgevoerd op het niveau van werkproces en omvat het vaststellen van relevante kwaliteitsdimensies. Deze zijn afhankelijk van het doel van het proces en de eisen van de relevante stakeholders. Van die dimensies wordt beoordeeld of deze in voldoende mate in het proces zijn geborgd. Tekortkomingen worden vastgelegd in requirements die zowel afzonderlijk beschreven en gemodelleerd worden als gerelateerd aan het procesmodel. Het realiseren van de requirements vindt plaats volgens de geldende procedures binnen de organisatie.

De methode is in één iteratie ontwikkeld en toegepast op een case, het werkproces controleren en opslaan van ingewonnen gegevens, waarbij de iteratie onderverdeeld is in drie fases. De methode is getoetst in bijeenkomsten met een klankbordgroep binnen de onderzoeksorganisatie en met een schriftelijke vragenlijst waarbij de methode op de criteria uit de literatuur getoetst is. De verzamelde feedback is geanalyseerd en verwerkt in een nieuwe versie van de methode die vervolgens opnieuw is toegepast op de case. De methode voldoet over het algemeen aan de gestelde criteria. Omdat er maar één iteratie uitgevoerd is, zijn er nog wel een aantal punten om mee te nemen in een vervolgonderzoek. Stap 4 kan verdiept worden met een simulatie om zo meer problemen boven tafel te krijgen, de aansluiting tussen verbetervoorstellen uit deze methode en de voortbrenging van de informatievoorziening (IV) kan doorontwikkeld worden, het activiteitsniveau in processen wordt meegenomen en de methode kan op een andere case in een andere context meegenomen worden. Daarmee wordt ook de externe validiteit vergroot.

## Summary

Both in practice and in literature, the importance of data as a key asset in organizations and the quality of data is recognized. Data quality is the extent to which data can be used for the given purpose in a given context. The quality of data can be assessed through dimensions. Dimensions provide metrics for measuring quality. Examples of quality dimensions are accuracy, credibility, completeness, accessibility and comprehensibility.

There are various situations of common data quality problems. The credibility of data is adversely affected by multiple sources for the same data. This ensures that the data is not used, reducing its added value. This leads to the core problem; no use of data. For operational problems in data production, changing needs of data buyers and distributed data processing, similar patterns are known that ultimately prevent data from being used. The causes of data quality problems are diverse. For example, dimensions can influence each other. Striving for timeliness can have consequences for completeness or accuracy. In addition, factors play a role in the information processing process. Furthermore, data quality problems are inevitable if systems are closely connected and the components of systems have complex interaction patterns.

Data quality problems can have a negative impact on business processes. Almost all methods to improve data quality contain three phases. A descriptive phase, an assessment phase in which the quality dimensions are assessed, and an improvement phase in which strategies and techniques are used to achieve a higher quality of data. The most important strategies for solving data quality issues are data-driven strategies and process-driven strategies. The process-oriented approach has two main approaches: controlling the process or redesigning the process. Both approaches ensure that problems are prevented and, from a long-term perspective, are preferable to the data-driven strategy.

The aim of this research is to provide organizations with tools to make data quality problems in business terms understandable, to analyze and to solve. The main question of this research is to investigate to what extent a data quality-oriented method of process modelling can help. Several methods have been developed that aim to achieve this, but to a limited extent meet the criteria set by stakeholders. These criteria can be subdivided into three sections:

- Principles, models must be consistent with standard Business Process Management (BPM) methods and must be familiar to stakeholders.
- General requirements, including that models must show processes that are necessary to produce an information product in a systematic way, that it can be used as input for a system development process and that boundaries of the organizations and systems involved must be visible.
- Requirements focused on data quality, including that quality controls must be visible, critical phases that influence quality must be transparent and that the model must be supportive to implement quality measures.

DQPM, Data Quality Process Modelling, is a seven-step method based on these criteria. Figure 2 shows the relationship between the steps.

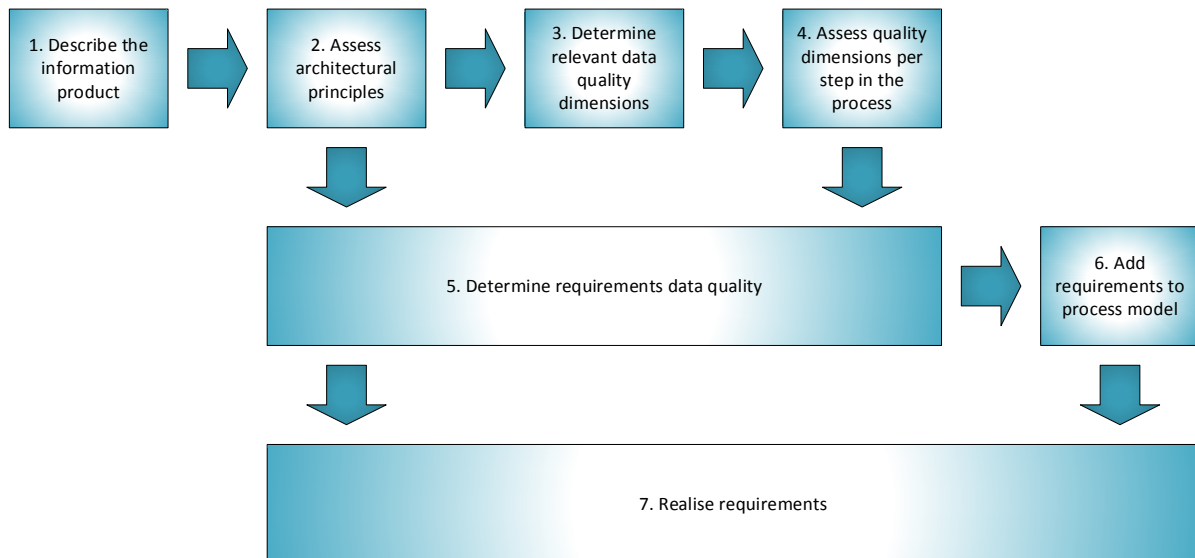


Figure 2 Overview DQPM

The first step describes the information product, the requirements that customers place on it and the place in the process where the product is created. This is done at work process level. In the second step, the entire business process of which the work process is a part, is examined and an assessment is made as to whether the model complies with important architectural principles such as "quality at source", determining the quality of data at the start of the process. Depending on the outcome, this results in requirements for adjusting the process. The next step is at the level of the work process and involves determining relevant quality dimensions. These depend on the purpose of the process and the requirements of the relevant stakeholders. For those dimensions, it is assessed whether they are sufficiently assured in the process. Shortcomings are recorded in requirements that are described and modelled separately as well as related to the process model. The requirements are realized in accordance with the applicable procedures within the organization.

The method was developed in one iteration and applied to a case, the work process 'Checking and storing collected data', whereby the iteration is subdivided into three phases. The method was tested in meetings with a focus group within the research organization and with a written questionnaire in which the method was tested on the criteria from the literature. The collected feedback was analysed and processed in a new version of the method that was subsequently applied to the case. The method generally meets the set criteria. Because only one iteration has been carried out, there are still a number of points to include in a follow-up study. Step 4 can be deepened with a simulation to get more problems out of the table, the connection between improvement proposals from this method and the production of information provision (IV) can be further developed, the level of activity in processes can be taken into account and the method can be taken on another case in a different context. This also increases external validity.



# Inhoudsopgave

|        |   |    |
|--------|---|----|
| 1.     | Introductie .....   | 10 |
| 1.1.   | Inleiding.....  | 10 |
| 1.2.   | Gebiedsverkenning .....                                     | 10 |
| 1.3.   | Probleemstelling .....                                      | 11 |
| 1.4.   | Onderzoeksvragen .....                                      | 11 |
| 1.5.   | Relevantie .....  | 12 |
| 1.6.   | Aanpak in hoofdlijnen .....                                 | 12 |
| 2.     | Theoretisch kader .....                                     | 13 |
| 2.1.   | Onderzoeksaanpak.....                                       | 13 |
| 2.2.   | Uitvoering.....   | 13 |
| 2.3.   | Resultaten literatuuronderzoek.....                         | 14 |
| 2.3.1. | Hoe wordt datakwaliteit bepaald? .....                      | 14 |
| 2.3.2. | Welke factoren beïnvloeden datakwaliteit?.....              | 16 |
| 2.3.3. | Welke methoden zijn er om datakwaliteit te verbeteren ..... | 19 |
| 2.3.4. | Datakwaliteit in procesmodellen .....                       | 20 |
| 2.3.5. | Eisen aan procesmodellen met datakwaliteit.....             | 21 |
| 2.4.   | Conclusies .....  | 22 |
| 2.5.   | Doel van het vervolgonderzoek .....                         | 23 |
| 3.     | Methodologie.....   | 24 |
| 3.1.   | Methode van onderzoek.....                                  | 24 |
| 3.2.   | Plan van aanpak .....                                       | 25 |
| 3.2.1. | Stap 3, Ontwerp en ontwikkeling.....                        | 25 |
| 3.2.2. | Stap 4, Demonstratie .....                                  | 25 |
| 3.2.3. | Stap 5, Evaluatie.....                                      | 26 |
| 3.2.4. | Stap 6, Communicatie .....                                  | 26 |
| 3.3.   | Methodologische issues.....                                 | 26 |
| 4.     | Resultaten .....  | 29 |
| 4.1.   | Selecteren modelleertaal.....                               | 29 |
| 4.2.   | Ontwerp van de methode DQPM .....                           | 29 |
| 4.2.1. | Overview van de methode.....                                | 29 |
| 4.2.2. | Onderbouwing van de methode .....                           | 30 |
| 4.2.3. | Uitgangspunten voor gebruik van de methode .....            | 31 |

|        |  |    |
|--------|--|----|
| 4.2.4. | Toelichting per stap .....                                       | 32 |
| 5.     | Demonstratie: toepassen DQPM .....                               | 42 |
| 5.1.   | Vorbereiding .....   | 42 |
| 5.1.1. | Selecteren case .....  | 42 |
| 5.1.2. | Selecteren klankbordgroep.....                                   | 42 |
| 5.1.3. | Vaststellen criteria .....                                       | 43 |
| 5.2.   | Toepassing op de case .....                                      | 44 |
| 5.3.   | Verzamelen feedback klankbordgroep .....                         | 54 |
| 5.3.1. | Feedback klankbordgroep.....                                     | 55 |
| 6.     | Evaluatie DQPM .....   | 57 |
| 7.     | Conclusie, discussie en aanbevelingen, reflectie .....           | 59 |
| 7.1.   | Discussie.....   | 59 |
| 7.2.   | Conclusies .....   | 60 |
| 7.3.   | Aanbevelingen voor de praktijk .....                             | 61 |
| 7.4.   | Aanbevelingen voor verder onderzoek.....                         | 62 |
| 7.5.   | Reflectie .....  | 63 |
|        | Referenties .....  | 64 |
|        | Bijlage 1 Dimensies van datakwaliteit .....                      | 67 |
|        | Bijlage 2 Concepten IP-MAP vs BPMN concepten .....               | 72 |
|        | Bijlage 3 Requirements aan procesmodellen en datakwaliteit ..... | 73 |
|        | Bijlage 4 Beslistabellen stap 4 DQPM .....                       | 75 |
|        | Bijlage 5 Verantwoording samenstelling criteria.....             | 78 |
|        | Bijlage 6 BPMN concepten gegevensgebruik.....                    | 81 |
|        | Bijlage 7 Feedback klankbordgroep .....                          | 82 |
|        | Bijlage 8 Vragenlijst klankbordgroep .....                       | 89 |
|        | Bijlage 9 Antwoorden vragenlijst klankbordgroep .....            | 94 |

# 1. Introductie

## 1.1. Inleiding

De Belastingdienst krijgt jaarlijks grote hoeveelheid gegevens aangeleverd van banken; gegevens over rekeninghouders en saldi van bankrekeningen. Deze gegevens worden voor diverse doeleinden gebruikt. Bij de aanlevering van vorig jaar bleek een van de banken bedragen in centen aangeleverd te hebben in plaats van in euro's. De systemen die de bankgegevens gebruiken, gaan uit van euro's. Deze vergissing werd tijdig ontdekt, maar als dit niet het geval was geweest zouden de gevolgen groot zijn voor zowel interne als externe afnemers: verkeerde berekeningen waar verkeerde conclusies aan verbonden zouden worden. Daarnaast zou een en ander breed uitgemeten worden in de pers.

Niet alleen in de praktijk, maar ook in de literatuur wordt onderkend dat het belang van data en het belang van de kwaliteit van data in organisaties groot is (Ofner, Otto, & Österle, 2012), vooral in informatie-intensieve organisaties (Cappiello et al., 2013). Data wordt gezien als een 'key asset' die helpt om sociale, economische en bedrijfskundige ontwikkelingen te begrijpen en om de concurrentiepositie van organisaties te verbeteren (Mezzanzanica, Boselli, Cesarini, & Mercorio, 2015). Nass en Scheibmayer (2011) leggen een relatie tussen de impact van datakwaliteit op bedrijfsprocessen door verstoringsvariabelen en de impact van deze variabelen op de ondernemingsresultaten. Volgens onderzoek hebben ondernemingen foutpercentages in hun systemen tot meer dan 20%, wat zorgt voor vertraging in orderafhandelingsprocessen en een negatieve invloed heeft op de voordelen van CRM-, ERP- en BI-systemen.

In dit afstudeerproject wordt gezocht naar een manier om problemen met datakwaliteit en hun oorzaken inzichtelijk te maken in bedrijfsprocessen.

In het vervolg van dit hoofdstuk wordt na een verdere gebiedsverkenning de probleemstelling, het doel en de onderzoeksvragen toegelicht.

## 1.2. Gebiedsverkenning

Een aantal begrippen die in dit rapport een rol spelen zullen hier kort toegelicht worden.

Datakwaliteit is de mate waarin data bruikbaar is voor het gegeven doel in een gegeven context (Yang et al., 2013). Datakwaliteit is een multidimensionaal concept dat zowel objectief te meten is als subjectief beleefd kan worden door medewerkers in organisaties (Pipino, Lee, Wang, Lowell Yang Lee, & Yang, 2002). Een datakwaliteitsdimensie is een aspect of eigenschap van informatie en een manier om informatie en behoeften ten aanzien van datakwaliteit te classificeren. Dimensies worden gebruikt om de kwaliteit van data te definiëren, te meten en te beheersen. Voorbeelden van dimensies zijn geloofwaardigheid, compleetheid en relevantie. Een dimensie kan betrekking hebben op de waarde van data of op de betekenis ervan in een logisch of conceptueel gegevensmodel.

Er zijn twee belangrijke strategieën om problemen met betrekking tot datakwaliteit in relatie tot een of meerdere van de genoemde dimensies op te lossen (Batini, Cappiello, Francalanci, & Maurino, 2009):

- Datagedreven strategieën;
- Procesgedreven strategieën.

Datagedreven strategieën richten zich op het manipuleren van data, tijdens de verwerking of achteraf na controles. Datagedreven strategieën beperken zich vaak tot juistheid van gegevens en zorgen veelal niet dat fouten voorkomen worden (Lee, Pipino, Funk, & Wang, 2006). De procesgedreven strategieën zien het voortbrengen van een informatieproduct als een manufacturing

proces waarin een product geproduceerd wordt en waarin procesverbeteringen aangebracht kunnen worden om de kwaliteit van het product te verbeteren (Shankaranarayanan, Wang, & Ziad, 2000). Hierbij kan Business Process Management (BPM) als kennisgebied helpen. BPM biedt methoden om processen te analyseren en te verbeteren (Van der Aalst, 2013).

### 1.3. Probleemstelling

Data is een kritieke factor voor organisaties om strategische en operationele doelstellingen te realiseren (Ofner et al., 2012). Door kwaliteitsproblemen met data lopen organisaties risico's. Slechte datakwaliteit heeft invloed op de efficiency van processen en de effectiviteit van beslissingen die op onbetrouwbare gegevens zijn gebaseerd (Cappiello et al., 2013).

Eén van de twee strategieën om de kwaliteit van data te verbeteren is het proces dat informatieproducten voortbrengt te verbeteren en te optimaliseren (Batini et al., 2009). Om inzichtelijk te maken hoe het proces moet veranderen is het nodig om in het proces die plaatsen te kunnen herkennen waar de kwaliteit van data beïnvloed wordt (Shankaranarayanan et al., 2000). Door vooraf processen te toetsen op deze kritieke plaatsen, kunnen tijdig maatregelen ontworpen worden om de kwaliteit van data te verhogen. Business Process Management (BPM) biedt een algemene structuur om processen te verbeteren en te analyseren (model, uitvoering, analyse, verbeteren) (Van der Aalst, 2013), maar er is nog weinig onderzoek gedaan naar de manieren waarop vanuit de optiek van BPM de kwaliteit van data verbeterd kan worden.

Dit onderzoek richt zich op hoe de beïnvloeding van datakwaliteit in het informatieverwerkende proces zichtbaar gemaakt kan worden om daarmee vanuit een procesperspectief de kwaliteit van data te verbeteren. Hiermee krijgen organisaties meer grip op de beheersing van hun datakwaliteit zodat daarmee de risico's voor organisaties beperkt worden, de efficiency van processen verhoogd wordt en de effectiviteit van beslissingen verbeterd.

### 1.4. Onderzoeksvragen

#### **Opdracht**

De opdracht is te onderzoeken hoe een datakwaliteitsgeoriënteerde manier van procesmodelleren stakeholders in een organisatie kan helpen om grip te krijgen op datakwaliteit om daarmee de kwaliteit van data te verhogen.

#### **Doel**

Het doel wat daarmee beoogd wordt is om organisaties handvatten te geven problemen met datakwaliteit, die vaak technisch van aard zijn, zichtbaar te maken, in business termen begrijpelijk te maken en te kunnen analyseren en op te kunnen lossen in zowel de ontwerpfase als de uitvoeringsfase van een proces.

#### **Centrale onderzoeksvraag**

In welke mate is een datakwaliteitgeoriënteerde manier van procesmodelleren bruikbaar voor stakeholders in een organisatie om de problemen met datakwaliteit op te sporen en inzichtelijk te maken?

#### **Afgeleide onderzoeksvragen**

1. Hoe wordt de kwaliteit van data bepaald?
2. Welke factoren beïnvloeden datakwaliteit?
3. Welke methoden zijn er om datakwaliteit te verbeteren?

4. Hoe kunnen factoren die datakwaliteit beïnvloeden zichtbaar gemaakt worden in een procesmodel?
5. Welke eisen stellen stakeholders aan een procesmodel dat laat zien waar datakwaliteit beïnvloed wordt?
6. In hoeverre helpt een datakwaliteitgeoriënteerde procesmodel om problemen met datakwaliteit op te sporen en inzichtelijk te maken?

## 1.5. Relevantie

Om de procesgedreven strategieën voor het verbeteren van datakwaliteit zoals Lee et al. (2006) en Batini et al. (2009) aanbevelen te ontwikkelen is het nodig inzicht te hebben in het proces dat informatieproducten voortbrengt. Een model van het proces helpt hierbij. Van der Aalst (2013) beschrijft vier belangrijke stappen in het beheersen en verbeteren van processen: modelleren van processen, uitvoeren waarbij het model gebruikt wordt om te beheersen en te ondersteunen, analyseren en vervolgens doorvoeren van beheersingsmaatregelen. Voor het modelleren van processen vanuit het perspectief van datakwaliteit bieden standaard modelleermethoden zoals BPMN en UML Activity Diagrams te weinig houvast (Ofner et al., 2012). Er zijn methoden zoals IP-Map beschreven door Shankaranarayanan et al. (2000) die herontwerp van processen als uitgangspunt nemen om datakwaliteit te verbeteren. Over het algemeen is er echter nog weinig onderzoek gedaan naar het beschrijven en verbeteren van datakwaliteit in de context van BPM (Ofner et al., 2012).

Het doel van dit onderzoek is om de kennis te vergroten op het gebied van de mogelijkheden om datakwaliteit te verbeteren vanuit een procesperspectief. Dit helpt organisaties om meer grip te krijgen op de kwaliteit van hun data. Data wordt steeds belangrijker als asset voor een organisatie. Het verbeteren van de kwaliteit van die data kan zorgen voor een betere concurrentiepositie (Mezzanzanica et al., 2015).

## 1.6. Aanpak in hoofdlijnen

Het volgende hoofdstuk beschrijft de resultaten van het literatuuronderzoek. Daarin wordt het onderzoeksterrein in kaart gebracht en wordt aangegeven welke methoden er zijn om datakwaliteit te verbeteren en welke daarvan vanuit procesperspectief ontwikkeld zijn. Aan de hand van criteria uit de literatuur wordt aangegeven wat de kwaliteit is van de bestaande methoden. De aanpak van het praktijkonderzoek wordt beschreven in hoofdstuk drie. Uit de literatuur wordt op basis van de criteria een methode of aanvulling op een bestaande methode ontwikkeld. Deze wordt vervolgens getoetst in de praktijk. Het laatste hoofdstuk bevat discussie, conclusies en aanbevelingen.

## 2. Theoretisch kader

In dit hoofdstuk wordt het theoretisch kader voor het onderzoek beschreven.

### 2.1. Onderzoeksaanpak

#### *Doel theoretisch kader*

Het doel van het theoretisch kader is om het onderzoeksgebied in kaart te brengen en te beschrijven aan de hand van beschikbare bronnen in de academische literatuur. Het kader geeft antwoord op de eerste vijf afgeleide onderzoeksvragen en op wat er ontbreekt aan de bestaande methoden. Dat vormt de basis voor de uitvoering van het onderzoek.

#### *Onderzoeksaanpak*

Twee bronnen waren voor de start van het onderzoek al beschikbaar, de Code voor Informatiekwaliteit en het boek *Journey to dataquality*. De aanpak van het onderzoek was als volgt.

- De al beschikbare bronnen doornemen om een beeld te krijgen van het onderzoeksgebied.
- Systematisch per onderzoeksvraag artikelen selecteren aan de hand van zoekcriteria zoals die in tabel 1 zijn weergegeven. De tabel bevat alleen de initiële zoekcriteria, in het zoekproces worden deze verfijnd aan de hand van de getoonde resultaten.

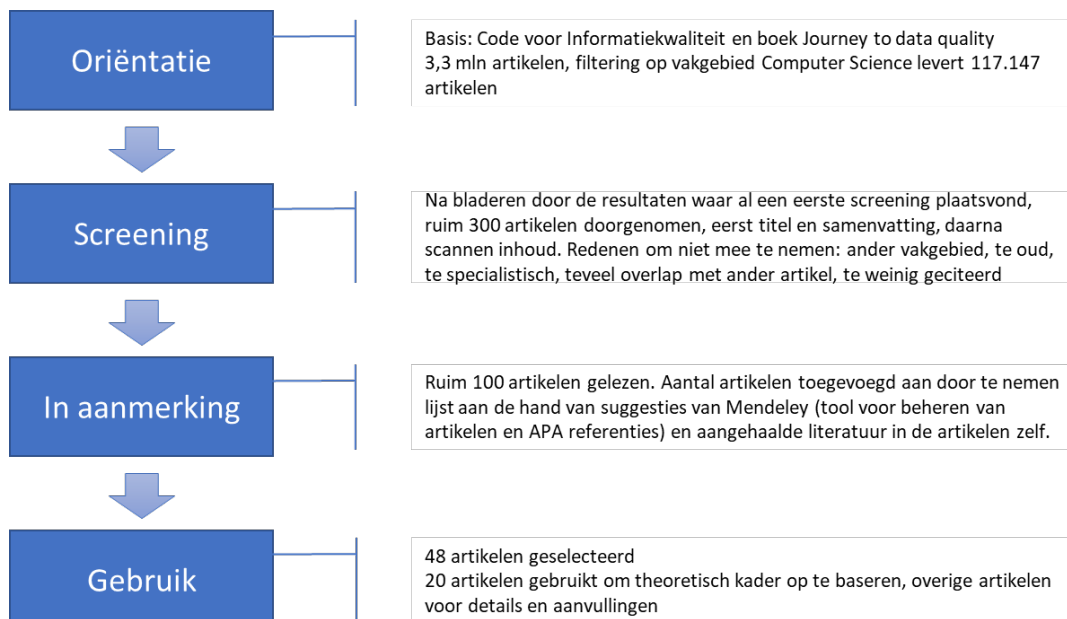
*Tabel 1 Onderzoeksvragen en zoektermen*

|   |   |
|---|---|
| 1. Hoe wordt datakwaliteit bepaald  | Data quality, data quality dimensions                                     |
| 2. Welke factoren beïnvloeden datakwaliteit   | Data quality problems, data quality issues                                |
| 3. Welke methoden zijn er om datakwaliteit te verbeteren?   | Data quality methodologies, data quality framework, data quality approach |
| 4. Hoe kunnen factoren die datakwaliteit beïnvloeden zichtbaar gemaakt worden in een procesmodel?         | Data quality process, data quality bpm, data quality bpmn                 |
| 5. Welke eisen stellen stakeholders aan een procesmodel dat laat zien waar datakwaliteit beïnvloed wordt? | Data quality process, data quality bpm                                    |

- De literatuurverwijzingen van de geselecteerde artikelen doornemen om aan de hand daarvan nog niet geraadpleegde bronnen te onderzoeken.

### 2.2. Uitvoering

De belangrijkste bron voor het onderzoek was de bibliotheek van de Open Universiteit. Door middel van de quick search functie is in alle beschikbare bronnen gezocht waarbij alleen de tijdschriftartikelen zijn gebruikt. De uitvoering verliep top-down zoals in figuur 2 weergegeven. Na een eerste oriëntatie zijn ruim 300 artikelen doorgenomen waarvan na screening zo'n 100 artikelen daadwerkelijk gelezen zijn. Er is in de selectie niet gefilterd op jaartal, bij de screening is wel naar de ouderdom van de artikelen gekeken. Daarvan zijn er 48 geselecteerd als basis voor het onderzoek. Twintig daarvan vormen de kern, de overige artikelen zijn gebruikt voor details of kleine aanvullingen op de andere artikelen.



Figuur 3 Uitvoering literatuuronderzoek

## 2.3. Resultaten literatuuronderzoek

Deze paragraaf bevat de resultaten van het literatuuronderzoek. Voor de eerste vijf onderzoeksvragen worden de antwoorden gegeven die gevonden zijn in de literatuur.

### 2.3.1. Hoe wordt datakwaliteit bepaald?

#### *Visies op datakwaliteit*

De literatuur geeft verschillende definities van datakwaliteit. Een algemene definitie is die van (Aljumaili, Karim, & Tretten, 2016; Lee, Strong, Kahn, & Wang, 2002): datakwaliteit is de mate waarin data geschikt is voor gebruik door gebruikers. Panwar (2015) maakt deze concreter door het gebruik toe te spitsen op het nemen van beslissingen: datakwaliteit is een essentiële eigenschap die de betrouwbaarheid bepaalt van data ten behoeve van het nemen van beslissingen (Panwar, 2015).

In plaats van gebruikers noemt McGilvray het gebruik van data voor toepassingen: datakwaliteit is de mate waarin gegevens een betrouwbare bron vormen voor één of meerdere toepassingen. De definitie die de voorgaande definities in meer of mindere mate omvat is die van (Yang et al., 2013): *datakwaliteit is de mate waarin data bruikbaar is voor het gegeven doel in een gegeven context*. Vanwege het omvattende karakter hanteren we voor dit onderzoek verder de definitie van Yang et al. De *context* waar Yang et al. naar verwijst is een belangrijk element in de definitie. Madnick en Hongwei Zhu (2006) noemen in hun onderzoek een voorbeeld, de price-earnings ratio<sup>1</sup> (P/E-ratio) voor DaimlerChrysler. Bron ABC heeft als waarde 11,6, Bloomberg 5,57, DBC 19,19 en Marketguide 7,46. De hoogste is bijna vier keer zo groot als de laagste, welke is correct? Het verrassende antwoord is dat ze alle vier correct zijn. Het is afhankelijk van hoe 'inkomsten' gedefinieerd wordt, inkomsten van het afgelopen kwartaal of over het afgelopen jaar. En welk jaar? Afgelopen kalenderjaar? De laatste twaalf maanden? Het laatste fiscale jaar? Wat op het eerste gezicht een datakwaliteitsprobleem lijkt, verschillende waardes voor dezelfde ratio, is feitelijk een interpretatieprobleem. Data uit de ene context kan niet zondermeer in de andere context gebruikt worden.

<sup>1</sup> Een ratio die iets zegt over de waarde van een onderneming, het laat de verhouding zien tussen de marktwaarde van een aandeel en de inkomsten van een aandeel.

Pipino et al. (2002) noemen datakwaliteit een multidimensionaal begrip: *subjectief* en *objectief*. Subjectief is de kwaliteit zoals die wordt ervaren door “degenen die ermee te maken hebben.” Objectief is de kwaliteit zoals die vastgesteld kan worden door het uitvoeren van metingen op dataverzamelingen. “Degenen die ermee te maken hebben” is een ruim begrip, (Lee & Strong, 2003) noemen drie rollen: *data collectors*, degenen die data verzamelen, *data custodians*, degenen die verantwoordelijk zijn voor opslag en beheer van data en *data consumers*, degenen die data gebruiken voor verdere integratie, aggregatie, presentatie en interpretatie.

De soort data is van invloed op de mate waarin kwaliteit beheerst kan worden. (Aljumaili et al., 2016; Batini et al., 2009) noemen de volgende indeling:

- *Gestructureerde data*; items beschreven in elementaire attributen die gedefinieerd zijn binnen een bepaald domein. Voorbeelden: data in relationele tabellen.
- *Ongestructureerde data*; een opsomming van tekens, veelal uitgedrukt in een natuurlijke taal.
- *Semigestructureerde data*; data waarvan de structuur een zekere mate van flexibiliteit heeft. Wordt ook wel data zonder schema of zelfbeschrijvend genoemd, bijvoorbeeld XML.

Naarmate data minder gestructureerd wordt, is het moeilijker om de kwaliteit van data te beheersen. De meeste datakwaliteitsmethoden beperken zich tot gestructureerde en semigestructureerde data. Methoden voor datakwaliteit die data zien als het voortbrengen van een product kennen een andere indeling:

- *Ruwe data*; data die geen bewerking heeft ondergaan na creatie en eerste opslag.
- *Informatieproducten*; het resultaat van een manufacturingactiviteit op data.

*Samengestelde data item*; wordt gegenereerd als een informatieproduct nodig is en slechts tijdelijk bewaard wordt totdat het eindproduct is samengesteld.

#### *Meetbaarheid van datakwaliteit*

Door middel van dimensies kan de kwaliteit van data beoordeeld worden. Dimensies geven metrieken om kwaliteit te kunnen meten (Batini et al., 2009). Wang, Reddy, en Kon (1995) onderkennen al een hiërarchie in kwaliteitsdimensies, maar die is nog vrij rudimentair met vier hoofdgroepen: toegankelijkheid, interpreteerbaarheid, bruikbaarheid en geloofwaardigheid.

Lee et al. (2002) hebben een meer uitgewerkte indeling gemaakt waarbij vier invalshoeken onderscheiden worden waarbij per invalshoek de dimensies weergegeven worden, zie tabel 2.

*Tabel 2 Overzicht kwaliteitsdimensies volgens Lee et al., 2002*

| Intrinsieke kwaliteit        | Invalshoek informatiekwiteit |                           |                    |
|------------------------------|------------------------------|---------------------------|--------------------|
|                              | Context                      | Representatie             | Toegankelijkheid   |
| <b>Juistheid<sup>2</sup></b> | Waarde toevoegend            | Begrijpelijkheid          | Toegankelijkheid   |
| <b>Geloofwaardigheid</b>     | Relevantie                   | Interpreteerbaarheid      | Gebruiksgemak      |
| <b>Reputatie</b>             | Volledigheid                 | Beknpte representatie     | Beveiliging        |
| <b>Objectiviteit</b>         | Tijdigheid                   | Consistente representatie | Lokaliseerbaarheid |
| <b>Feitelijkheid</b>         | Kwantiteit                   | Leesbaarheid              |                    |
| <b>Consistentie</b>          | Betrouwbaar                  | Syntax                    |                    |

<sup>2</sup> Hier is ‘juistheid’ gekozen als vertaling voor de Engelse term ‘accuracy’



|                        |               |                          |  |
|------------------------|---------------|--------------------------|--|
| <b>Compleetheid</b>    | Gebruik       | Semantische consistentie |  |
| <b>Precisie</b>        | Bruikbaarheid | Aliases                  |  |
| <b>Betrouwbaarheid</b> |               | Oorsprong                |  |
| <b>Vrij van bias</b>   |               | Duidelijkheid            |  |
| <b>Correctheid</b>     |               | Beknoptheid              |  |
|                        |               | Uniekheid                |  |
|                        |               | Vergelijkbaarheid        |  |
|                        |               | Compatibiliteit          |  |
|                        |               | Betekenis                |  |

Opvallend is dat wat Wang et al. (1995) een hoofdrubriek noemt, hier een niveau dieper geplaatst wordt als dimensie. Toegankelijkheid is in beide onderzoeken een hoofdrubriek. Bruikbaarheid is bij Wang et al. (1995) een hoofdrubriek en bij Lee et al. (2002) slechts een element. Kijken we naar de definitie die in dit onderzoek gebruikt wordt, “datakwaliteit is de mate waarin data bruikbaar is voor het gegeven doel in een gegeven context” (Yang et al., 2013), dan sluit de indeling van Lee et al. (2002) hier goed bij aan. Contextuele kwaliteit is een aparte hoofdrubriek waarin naast bruikbaarheid ook andere dimensies genoemd worden die in een bepaalde context voor een bepaald doel relevant zijn. Aan tijdigheid bijvoorbeeld kan door handelaren op de beursvloer, waar seconden soms van belang kunnen zijn, andere eisen gesteld worden dan door het management van dezelfde organisatie die wekelijks een overzicht wil hebben van de gerealiseerde transacties. Ook toegankelijkheid is een invalshoek die aan de ene kant iets zegt over de data zelf, maar ook over hoe de afnemer, de data consumer zoals Lee en Strong (2003) die noemen, de gegevens kan gebruiken. Ook representatie heeft dit duale aspect in zich. Alleen intrinsieke kwaliteit is gericht op de data zelf. Lee et al. (2002) nemen om deze reden nog een tweede tabel op die niet vanuit de academische invalshoek is opgesteld, maar vanuit de invalshoek van de praktijk. Deze bevat echter minder aanknopingspunten om de dimensies verder te beschrijven.

In veel artikelen en methoden voor het beheersen van datakwaliteit wordt een subset van de dimensies gebruikt. In bijlage 1 worden de dimensies weergegeven met hun definitie zoals die in dit onderzoek verder gebruikt gaan worden. De indeling is gebaseerd op de academische rubricering in Tabel 2 *Overzicht kwaliteitsdimensies volgens Lee et al., 2002*, de selectie is gebaseerd op de meest voorkomende dimensies in de gebruikte literatuur.

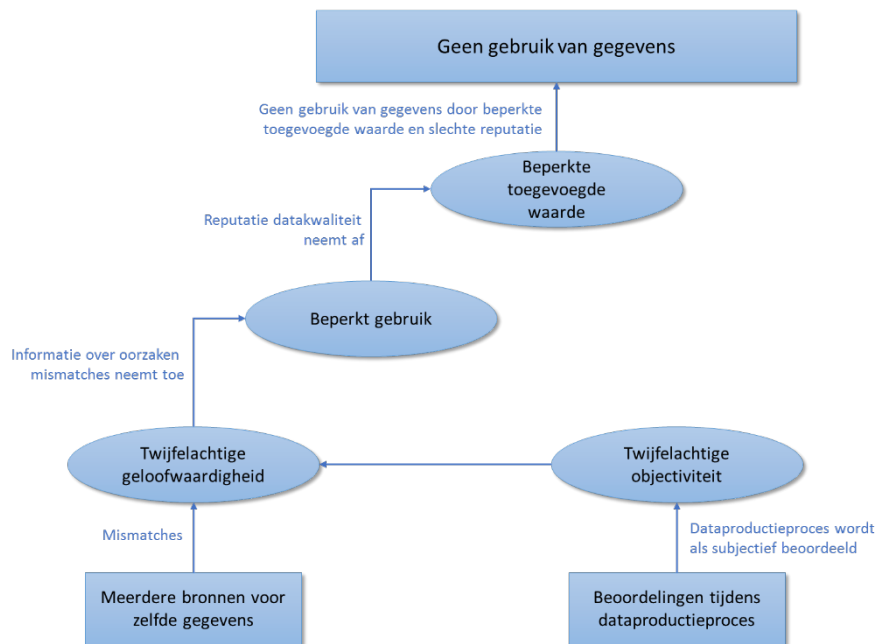
### 2.3.2. Welke factoren beïnvloeden datakwaliteit?

#### *Veel voorkomende problemen*

Het beheersen van datakwaliteit is moeilijk, de kwaliteit is afhankelijk van het proces als geheel en alle betrokkenen die daarin een rol spelen (Storey, Dewan, & Freimer, 2012). Strong, Lee, en Wang (1997) beschrijven in hun onderzoek naar veel voorkomende problemen een aantal patronen per invalshoek.

#### **Intrinsieke kwaliteit**

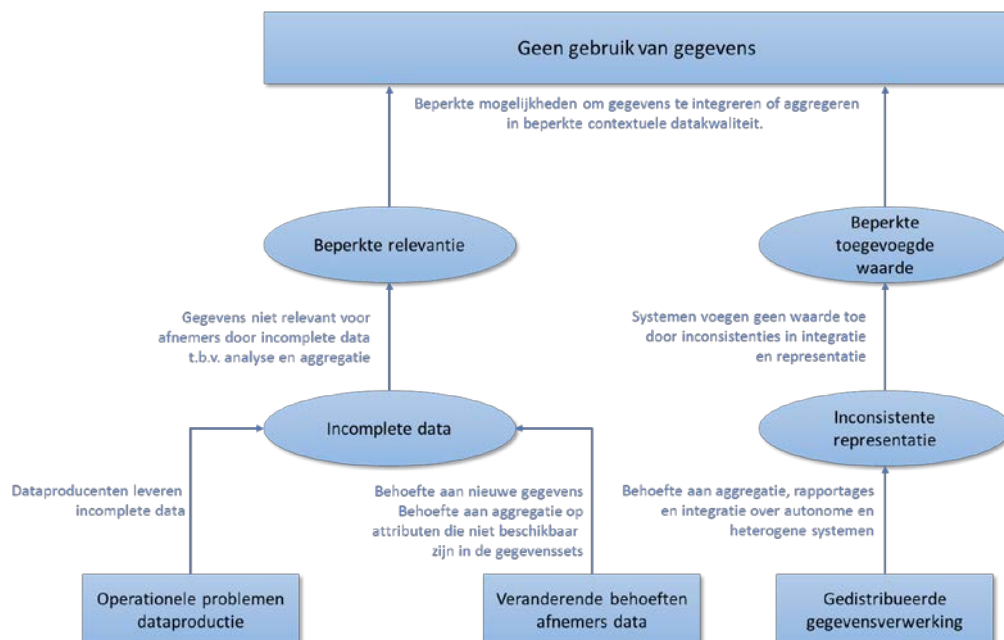
Door meerdere bronnen voor dezelfde gegevens wordt de geloofwaardigheid van data nadelig beïnvloed. Ook twijfelachtige objectiviteit, veroorzaakt door subjectieve beoordelingen in het dataproductieproces, dragen hieraan bij. Twijfelachtige geloofwaardigheid leidt tot minder gebruik van gegevens waardoor de toegevoegde waarde afneemt. Dit leidt uiteindelijk tot het kernprobleem, geen gebruik van gegevens. Zie Figuur 4 waarin het door Strong et al. (1997) beschreven patroon is weergegeven.



Figuur 4 Patroon problemen intrinsieke kwaliteit (Strong et al., 1997)

## Context

Ook voor contextuele kwaliteit onderkennen Strong et al. (1997) een patroon dat ervoor zorgt dat het gebruik van gegevens problemen oplevert. Problemen in het dataproductieproces en veranderende behoeften van afnemers van gegevens zorgen voor incomplete data. Dat leidt tot een beperkte relevantie voor gebruikers. Daarnaast zorgt een gedistribueerde verwerking voor inconsistente representatie van gegevens waardoor ook de toegevoegde waarde afneemt. Zie Figuur 5 waarin Strong et al. (1997) het patroon laten zien.



Figuur 5 Patroon problemen contextuele kwaliteit (Strong et al., 1997)

## Representatie en toegankelijkheid

Toegankelijkheid wordt in het onderzoek van Strong et al. (1997) beïnvloed door gebrek aan verwerkingscapaciteit waardoor systemen niet goed of beperkt toegankelijk zijn. Ook

beveiligingsmaatregelen om privacy te borgen hebben bedoeld maar ook onbedoeld effect op toegankelijkheid.

Interpreteerbaarheid en begrijpelijkheid worden negatief beïnvloed door interpretaties tijdens het verwerkingsproces van gegevens. Deze zorgen op hun beurt voor een minder goede toegankelijkheid. De hoeveelheid te verwerken gegevens kan een negatieve impact hebben op de tijdigheid van gegevens waardoor data ook slecht toegankelijk wordt.

#### *Oorzaken van problemen met datakwaliteit*

De oorzaken van problemen met datakwaliteit zijn divers. Onderzoek geeft verschillende oorzaken die hieronder besproken zullen worden.

#### **Wederzijdse beïnvloeding van dimensies**

Het onderzoek van Strong et al. (1997) laat een samenhang zien tussen verschillende dimensies. Problemen met tijdigheid hebben gevolgen voor de toegankelijkheid. Een beperkte semantische consistentie heeft gevolgen voor geloofwaardigheid en de mate van waarde die toegevoegd wordt. Dit zijn voorbeelden van dimensies die elkaar versterken, dat wil zeggen een beperking van de ene dimensie zorgt voor een beperking van de andere dimensies. Dimensies die elkaar negatief beïnvloeden hebben een negatieve invloed op datakwaliteit. Bijvoorbeeld het streven naar de verhoging van tijdigheid kan negatieve gevolgen hebben voor de volledigheid of juistheid van gegevens. Ook kan het streven naar volledigheid negatieve gevolgen hebben voor consistentie (Cao & Zhu, 2013).

#### **Factoren in het informatieverwerkende proces**

Veel onderzoekers bekijken het proces van voortbrengen van data door de bril van een manufacturingproces en stellen dat deze manier van kijken en daar naar handelen een effectieve manier is om datakwaliteit te beheersen (Batini et al., 2009; Lee et al., 2006; Strong et al., 1997). In de manufacturingwereld is lean een veelgebruikte manier om de kwaliteit van het proces en het product te verbeteren (Wahab, Mukhtar, & Sulaiman, 2013; Wee & Wu, 2009). Wahab et al. (2013) beschrijven de relatie tussen leandimensies in een manufacturingsysteem en verspilling. Ondanks de overeenkomst die gezien wordt tussen het voortbrengen van data en een manufacturingproces, is er in dit onderzoek geen publicatie te vinden die ingaat op het lean voortbrengen van data en de impact daarvan op de kwaliteit. Kijkend naar de kwaliteitsdimensies lijken er wel verspillingen uit de leanaanpak te zijn die gerelateerd kunnen worden aan datakwaliteitsdimensies. De manier van het inrichten van proces waarin data voortgebracht wordt, zou daarmee van invloed kunnen zijn op de kwaliteit van de data. Om dit verband aan te tonen is verder onderzoek nodig.

#### **Normal Accident Theory**

Cao & Zhu (2013) hebben onderzoek gedaan naar datakwaliteitproblemen die ontstaan als gevolg van de implementatie van ERP-systemen. Zij stellen dat datakwaliteitproblemen onoverkomelijk zijn als systemen nauw verbonden zijn en de componenten van de systemen complexe, verborgen interactiepatronen kennen. Deze stelling is gebaseerd op de Normal Accident Theory (NAT). Het onderzoek van Cao & Zhu (2013) liet zien dat problemen zich voordoen op de drie gebieden die Lee & Strong (2003) onderkennen: dataproducenten, -afnemers en -verzamelaars en dat de problemen verschillende dimensies van kwaliteit raken.

### 2.3.3. Welke methoden zijn er om datakwaliteit te verbeteren

Het inventariserend onderzoek van Batini et al., 2009 beschrijft meer dan tien methoden die gericht zijn op het in kaart brengen en verbeteren van datakwaliteit. Een aantal belangrijke (veel geciteerde) daarvan zijn:

- TDQM, Total Data Quality Management, Wang, 1998.
- AIMQ, A methodology for information quality assessment, Lee et al. 2002
- DQA, Data Quality Assessment, Pipino et al. 2002
- CDQ, Comprehensive methodology for Data Quality Management.

De lijst is niet uitputtend. Er zijn onderzoekers die bepaalde accenten leggen die niet in de opsomming van Batini et al., 2009 voorkomen. Madnick & Hongwei Zhu (2006) stellen bijvoorbeeld semantiek centraal en hebben een methode ontwikkeld, Context Interchange (COIN), om semantiek vast te leggen en semantische heterogeniteit in data in overeenstemming met elkaar te brengen. Ook zijn er onderzoekers die het gebruik van bestaande kwaliteitsmethodieken onderzocht hebben om datakwaliteit te verbeteren. Panwar (2015) heeft bijvoorbeeld onderzocht dat het gebruik van Six-Sigma effectief kan zijn in bepaalde situaties.

Bijna alle methoden die Batini onderzocht heeft bevatten drie fases:

1. Beschrijvende fase, waarin contextinformatie, processen, services, dataverzamelingen en kwaliteitsissues verzameld worden.
2. Assessment, waarin metingen worden gedaan voor verschillende kwaliteitsdimensies en deze worden beoordeeld om een beeld te vormen van de kwaliteit. De beschikbaarheid van metadata is belangrijk in deze fase omdat deze helpt de betekenis van data in een specifieke context te begrijpen (Aljumaili et al., 2016). In deze fase wordt ook een procesmodel gemaakt dat het produceren van data in beeld brengt. Lee, Pipino, Funk, en Wang (2009) hebben een vragenlijst ontwikkeld om naast de objectieve waarnemingen ook de beleving van kwaliteit door afnemers te meten.
3. Verbetering, waarin stappen, strategieën en technieken ingezet worden om een hogere kwaliteit van data te bereiken. Hier maken Batini et al. (2009) onderscheid tussen data- en procesgedreven verbeteringen. Onder datagedreven strategieën vallen bijvoorbeeld het verwerven van nieuwe -betere- data, standaardisatie, record linkage en data- en schema-integratie. De procesgedreven strategie kent twee hoofdbenaderingen:
  - a. *Beheersing van het proces* door het inbouwen van controleprocedures op het moment dat data gecreëerd wordt, gewijzigd wordt of benaderd wordt. Dit is een reactieve benadering die ervoor zorgt dat data niet in kwaliteit achteruit gaat en dat er geen onjuiste gegevens verspreid worden.
  - b. *Herontwerp van het proces* waarmee de oorzaken van slechte datakwaliteit weggenomen worden en waarmee nieuwe activiteiten toegevoegd kunnen worden in het proces om data van betere kwaliteit te produceren.

Opvallend is dat van de dertien onderzochte methoden er slechts zes de procesbenadering uitwerken, terwijl datagedreven strategieën vaak alleen op de korte termijn effectief zijn (Batini et al., 2009).

Başkarada en Koronios (2014) noemen een aantal randvoorwaarden om kwaliteit te verbeteren. Zij benaderen methoden om datakwaliteit te verbeteren vanuit het concept 'kritieke succesfactoren'. Wil een initiatief om datakwaliteit te verbeteren slagen, dan is betrokkenheid en steun van het hogere management belangrijk. Het plan en de doelstellingen moeten duidelijk zijn. Continu verbeteren kan bereikt worden door projecten van elkaar te laten leren.

Naast de wetenschappelijk onderzochte methoden zijn er nog internationale kaders die houvast geven om datakwaliteit te verbeteren. In de Code voor Informatiekwaliteit (2016) worden deze als volgt genoemd. “ISO 9001 (2014) is erop gericht om stelselmatige grip op kwaliteit aantoonbaar te realiseren. Deze normen beschrijven de karakteristieken van een systeem dat de kwaliteit van producten/diensten (dus ook van ‘informatie’) borgt en dat die borging ook zichtbaar maakt. Die zichtbaarheid impliceert veel aandacht voor vastleggingen en rapportages. ISO 8000 is een verzameling kaders en richtlijnen, gericht op de borging van kwaliteit van masterdata; in eerste instantie in een industriële context, waar data niet je product is.” (Vieveen et al., 2016).

#### 2.3.4. Datakwaliteit in procesmodellen

Ofner et al. (2012) geven een overzicht van methoden waarin datakwaliteit in processen weergegeven wordt. Dat begint met Ballou en Pazer die in 1985 algoritmes ontwikkelden om de impact van proces- en gegevenskwaliteit te onderzoeken. Deze aanpak had beperkingen waar in het Information Manufacturing System (IMS) aan tegemoet gekomen is. Het uitgangspunt veranderde door informatie als een product te zien dat geproduceerd wordt. In IMS is het mogelijk om attributen en gegevenselementen te volgen van bron tot eindproduct. IP-Map van Shankaranarayan et al. (2000) borduurde voort op dit idee en voegde meerdere concepten toe. Het ontwerp van een IP-Map gaat uit van de eisen die gesteld worden aan het eindproduct en laat zien welke gegevens en bronnen nodig zijn. Scannapieco (2006) maakte daar een UML-variant van. Ofner et al. (2012) stelde als tekortkoming vast dat in deze methoden elementaire BPM-concepten als procesbeheersing en doelen ontbraken. Zij sluiten daarom aan bij standaarden van de Object Management Group (OMG) en gebruiken Business Modeling and Management Specifications (BMMS), Semantics of Business Vocabulary and Business Rules (SBVR) en Business Process Model and Notation (BPMN) waarin deze concepten wel aanwezig zijn. Business rules spelen een belangrijke rol in het onderzoek van Ofner et al. (2012). De kwaliteit van de output van een processtap wordt bepaald door de formule  $DQ = 1 - E$ , waarbij E staat voor het percentage ‘data defects’. Businessregels bepalen of de output goed is of niet. Door de uitkomsten van de processtappen te sommeren wordt een beeld verkregen van de datakwaliteit van het hele proces.

Het verschil met IP-MAP is dat Ofner et al. (2012) gebruik maakt van een standaard modelleertaal. De concepten die in IP-MAP gebruikt worden, zijn stuk voor stuk te vertalen naar een BPMN concept (zie analyse in bijlage 2), alleen de manier waarop data gecontroleerd wordt is anders weergegeven. In de aanpak van Ofner is die vervat in regels die niet direct in het procesmodel terug te zien zijn. Die worden uitgevoerd in een bepaalde stap. In IP-MAP is een apart concept voor het controleren van data aan de hand van criteria. Daarmee voldoet IP-MAP aan een belangrijk requirement dat gesteld werd toen de methode ontwikkeld werd; het zichtbaar maken van de kritieke punten in het proces die de datakwaliteit beïnvloeden.

Glowalla & Sunyaev (2014) hebben ook een inventariserend onderzoek gedaan naar wat zij noemen PDDQM, proces gedreven datakwaliteit. Zij gaan dieper in op de inhoud van procesgedreven datakwaliteitmethoden. In hun inventarisatie staan de volgende toevoegingen genoemd met betrekking tot datakwaliteit in bestaande modelleertalen:

- Wat het meest voorkomt zijn controles op datakwaliteit, in aparte processtappen, als label, als element toegevoegd aan processtappen of als swimlane. In IP-Map is er een apart concept voor beschikbaar.
- Metrieken die een beeld geven van de datakwaliteit in een proces. De hiervoor genoemde aanpak van Ofner is een voorbeeld hiervan.

- Eén methode focust op het zichtbaar maken van tijdigheid in het proces door middel van het toevoegen van timestamps aan processtappen.
- Mielke (2005) heeft een eigen aanpak ontwikkeld waarin datakwaliteit over meerdere modellen heen geïntegreerd wordt. Op een globaal niveau wordt inzicht gegeven in de belangrijkste kwaliteitsdimensies voor de belangrijkste processen en afdelingen. Op een gedetailleerder niveau worden subprocessen en de in- en output voor informatieproducten weergegeven met een beeld van de datakwaliteit van het subprocess. De aanpak bestaat uit zeven stappen:
  1. Inventarisatie bij stakeholders wat de belangrijkste datakwaliteitsdimensies zijn
  2. Analyseren van bedrijfsprocessen en datastromen
  3. Met betrokken gebruikers kritieke punten in het proces vaststellen
  4. Opnieuw inventariseren bij stakeholders, inzoomen op datakwaliteit in proces, subprocess en kritieke punten
  5. Vaststellen metrieken om vast te stellen of voldaan wordt aan requirements van stakeholders
  6. Meten en analyseren
  7. Inventariseren bij stakeholders, herontwerp metrieken indien nodig en voer stap 1 opnieuw uit

Ondanks deze toevoegingen en de toenemende aandacht in de praktijk voor procesmodellen met een datakwaliteitstoevoeging, is er nog weinig onderzoek gedaan naar het datakwaliteitsperspectief in procesmodellering (Glowalla & Sunyaev, 2014).

### 2.3.5. Eisen aan procesmodellen met datakwaliteit

Verschillende onderzoeken hebben requirements geïnventariseerd die te stellen zijn aan een procesmodel dat datakwaliteit laat zien; Shankaranarayanan et al. (2000), Ofner et al. (2012) en Glowalla & Sunyaev (2014). De eisen uit de afzonderlijke onderzoeken zijn opgenomen in bijlage 3. In deze paragraaf zijn de eisen gerubriceerd en gecombineerd. Achter iedere eis staat de bron genoemd.

#### **Uitgangspunten**

- Modellen moeten aansluiten bij standaard BPM-methoden. *Ofner et al. (2012)*
- Modellen moeten vertrouwd zijn voor stakeholders, begrijpelijk zijn en niet te complex. *Glowalla & Sunyaev (2014)*

#### **Algemene eisen, niet specifiek voor datakwaliteit**

- De methode moet op een systematische manier de processen laten zien die nodig zijn om een informatieproduct voort te brengen. *Shankaranarayanan et al. (2000)*
- Het model moet inzicht geven in bottlenecks in het proces. *Shankaranarayanan et al. (2000)*
- Het model moet duidelijk maken wie eigenaar is van de processen die informatieproducten voortbrengen. *Shankaranarayanan et al. (2000)*
- De opgeleverde procesmodellen moeten bruikbaar zijn als input voor het systeemontwikkelingstraject. *Ofner et al. (2012)*
- Het model moet de raw dataflow, component dataflow en IP-flow kunnen laten zien. *Glowalla & Sunyaev (2014)*
- Het model moet de grenzen van betrokken organisaties, onderdelen daarvan en systemen laten zien. *Glowalla & Sunyaev (2014)*
- Het transformatieproces moet duidelijk zijn. *Glowalla & Sunyaev (2014)*

### Eisen gericht op datakwaliteit in procesmodellen

- De belangrijkste fasen in het proces en kritieke fasen die datakwaliteit beïnvloeden moeten gevisualiseerd kunnen worden. *Shankaranarayanan et al. (2000)*
- Kwaliteitscontroles zijn expliciet zichtbaar. *Glowalla & Sunyaev (2014)*
- Het model moet inzicht geven in de tijd die nodig is om een informatieproduct voort te brengen. *Shankaranarayanan et al. (2000)*
- Het model moet ondersteunend zijn in het implementeren van kwaliteitsmaatregelen. *Shankaranarayanan et al. (2000)*
- Het model moet het mogelijk maken om op verschillende abstractieniveaus de kwaliteit van het informatieproduct meetbaar te maken aan de hand van de relevante kwaliteitsdimensies. *Ofner et al. (2012)*  
*Shankaranarayanan et al. (2000)*
- De methode moet concepten bevatten om het belang van de business impact en waarde van datakwaliteit te kunnen laten zien. *Ofner et al. (2012)*
- De methode moet rekening houden met de relevante context van data zodat een gedifferentieerde representatie van data mogelijk is. *Ofner et al. (2012)*

Kijken we naar de onderzochte methoden in de drie genoemde onderzoeken, dan scoort geen van de methoden goed op alle criteria. IP-Map bijvoorbeeld is geen standaard BPM-methode en mist onderdelen die wel in standaard BPM-methode opgenomen zijn. Dat uitgangspunt heeft Ofner et al. (2012) wel meegenomen in de door hen ontwikkelde methode, maar die laat niet zien welke fase datakwaliteit beïnvloeden en wat de oorzaak is van slechte kwaliteit. Veel van de andere onderzochte methoden kennen maar een beperkt aantal kenmerken, bijvoorbeeld alleen kwaliteitscontroles of alleen tijdigheid (Glowalla & Sunyaev, 2014).

## 2.4. Conclusies

Datakwaliteit is de mate waarin data bruikbaar is voor het gegeven doel in een gegeven context. Door middel van kwaliteitsdimensies als volledigheid, juistheid, geloofwaardigheid en bruikbaarheid is datakwaliteit meetbaar te maken. Er zijn verschillende oorzaken waardoor problemen met datakwaliteit kunnen ontstaan. In de aanpak om kwaliteit te verbeteren zijn twee benaderingen, datagedreven strategieën en procesgedreven strategieën. Die laatste aanpak geniet de voorkeur omdat die de oorzaken tegengaat. In de procesgedreven aanpak is de afgelopen decennia een bepaalde ontwikkeling te zien van algoritmes om de impact van proces- en gegevenskwaliteit te onderzoeken naar een verrijking van bestaande modelleertalen met datakwaliteitselementen waarbij een informatieproduct gezien wordt als een product dat geproduceerd wordt. Verschillende onderzoekers hebben criteria verzameld waar deze procesmodellen aan moeten voldoen. Deze zijn onder te verdelen in uitgangspunten, algemene – BPM-gerelateerde – eisen en eisen gericht op datakwaliteit in procesmodellen. De methoden van modelleren uit de literatuur voldoen in beperkte mate aan de criteria. Of een methode geeft een beeld van de datakwaliteit van een stap of een proces, maar laat niet de oorzaken zien of een methode laat alleen kwaliteitscontroles zien. Of een methode laat wel – beperkt – zien waar datakwaliteit beïnvloed wordt, maar voldoet niet aan algemene BPM-vereisten. Een doorontwikkelde methode die aan meer criteria voldoet kan helpen om organisaties meer grip te geven datakwaliteit.



## 2.5. Doel van het vervolgonderzoek

Het doel van het vervolgonderzoek is om een bestaande procesmodelleertaal die aansluit bij bestaande BPM-methoden zodanig door te ontwikkelen dat deze aan de belangrijkste eisen voldoet die gesteld worden aan datakwaliteit in procesmodellen. Door de toepasbaarheid van de doorontwikkelde taal te toetsen in de praktijk wordt een bijdrage geleverd aan het beantwoorden van de centrale onderzoeksvraag; het bepalen in welke mate een datakwaliteitsgeoriënteerde manier van procesmodelleren bruikbaar voor is stakeholders in een organisatie om de problemen met datakwaliteit op te sporen en inzichtelijk te maken.

Zoals uit het theoretisch kader blijkt, is het hiervoor nodig een modelleertaal te vinden die zowel vertrouwd is bij de stakeholders in de onderzoeksorganisatie als aansluit bij bestaande BPM-methoden. Verder is inzicht nodig in de eisen die betrokken stakeholders stellen aan datakwaliteit in procesmodellen. Voldoen de criteria die verzameld zijn in het theoretisch kader of worden er andere aanvullende eisen gesteld? Als met deze inzichten een manier ontwikkeld is om een modelleertaal te ontwikkelen dan moet deze getoetst worden in de praktijk op bruikbaarheid.



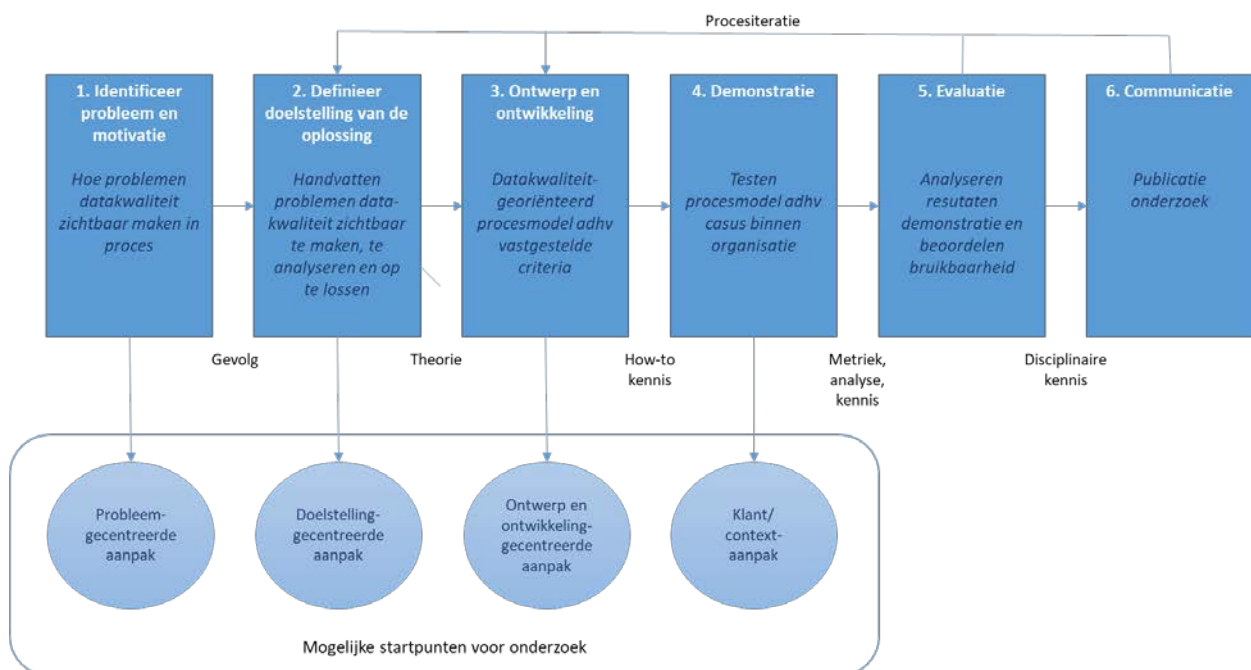
### 3. Methodologie

Dit hoofdstuk beschrijft de methode van onderzoek voor het empirische deel van het onderzoek.

#### 3.1. Methode van onderzoek

Voor het tweede deel van het onderzoek wordt gebruik gemaakt van Design Science Research Methodology (DSRM), een methode die onder andere ontwikkeld is door Peffers, Tuunanen, Rothenberger, en Chatterjee (2007). Peffers et al. (2007) stellen dat veel onderzoek in ICT is gericht op het begrijpen en verklaren van fenomenen zoals dat in natuurwetenschappen gebruikelijk is, maar dat die werkwijze minder geschikt is voor situaties waarin problemen opgelost moeten worden. Ook Hevner, March, Park, & Ram (2004) betogen dat design research een van de twee kenmerkende paradigma's is in ICT-onderzoek naast gedragswetenschap. Design research, waarbij een bruikbare oplossing voor een probleem wordt gecreëerd, is in andere disciplines zoals engineering al langer gebruikelijk, maar in ICT nog niet. Peffers et al. (2007) geeft een aantal regels waaraan design research moet voldoen; het onderzoek moet een product opleveren dat een probleem oplost, een nog niet of niet goed opgelost en belangrijk bedrijfsprobleem. De kwaliteit en bruikbaarheid moet grondig geëvalueerd worden. Het ontwikkelde product moet gecommuniceerd worden met de betrokken stakeholders. Zowel het doel van Design Research als de genoemde voorwaarden zijn van toepassing voor dit onderzoek. Het onderzoek dat Ofner et al. (2012) uitgevoerd heeft, heeft dezelfde doelstelling als dit onderzoek. Ook in dat onderzoek is gebruik gemaakt van design research waarbij de fasering dezelfde elementen kent als in DSRM, alleen teruggebracht tot vier fasen.

Voor dit onderzoek wordt de fasering van Peffers et al. (2007) gebruikt die schematisch als volgt weer te geven is. Cursief in de stappen is de uitwerking naar dit onderzoek opgenomen.



Figuur 6 DSRM stappen Peffers et al. (2007)

In paragraaf 3.2 worden deze stappen verder toegelicht.

## 3.2. Plan van aanpak

Dit onderzoek volgt de zes stappen van DSRM zoals in figuur 6 zijn opgenomen. Stap 1 en 2 zijn in hoofdstuk 1 van dit verslag uitgewerkt. Stap 3 en verder worden in het praktijkonderzoek uitgevoerd en worden hieronder verder toegelicht.

### 3.2.1. Stap 3, Ontwerp en ontwikkeling

Deze stap bestaat uit de volgende onderdelen:

a. Selecteren case

De case wordt geselecteerd in de onderzoeksorganisatie, en moet aan de volgende voorwaarden voldoen.

- het proces moet meerdere gegevensbewerkingen bevatten;
- het proces moet in ontwikkeling zijn;
- er moeten verschillende stakeholders betrokken zijn bij dit proces;
- het proces heeft bij voorkeur problemen met datakwaliteit.

De case zal worden gezocht binnen het domein 'Inwinnen en verstrekken van gegevens' binnen de onderzoeksorganisatie.

- b. Selecteren klankbordgroep. De klankbordgroep speelt een rol bij het toetsen van de criteria waar het model aan moet voldoen en bij stap 4, demonstratie. De klankbordgroep zal bestaan uit betrokkenen bij de uitvoering van het proces, functioneel beheerders, medewerkers ontwerp en automatisering en deskundigen op het gebied van methoden, technieken, hulpmiddelen en voorschriften.
- c. Vaststellen criteria waar de procesmodelleertaal die gebruikt gaat worden aan moet voldoen. De criteria uit de literatuur worden getoetst bij de klankbordgroep en vastgesteld. Hiervoor zullen gestructureerde en semigestructureerde interviews gebruikt worden als bron van primaire gegevens.
- d. Beoordelen huidig model van geselecteerde case op criteria datakwaliteit uit stap c.
- e. Selecteren van een procesmodelleertaal die geschikt is voor het onderzoek. De secundaire onderzoeksdata die nodig zijn voor deze stap bestaan in eerste instantie uit een overzicht van modelleertalen die aansluiten bij bestaande BPM-methoden en een beschrijving van de modelleermethoden die in de onderzoeksorganisatie gebruikt worden. Uit de primaire onderzoeksdata kan vastgesteld worden welke daarvan vertrouwd is bij de stakeholders in de onderzoeksorganisatie zodat een taal geselecteerd kan worden.
- f. Ontwikkelen van methode waarmee aan het doel van het onderzoek beantwoord wordt en die voldoet aan de criteria. Dit zal een iteratief en creatief proces zijn gebaseerd op de bevindingen uit het literatuuronderzoek en de vastgestelde criteria.

### 3.2.2. Stap 4, Demonstratie

In deze stap wordt de ontwikkelde methode getoetst in de praktijk. Binnen het domein waarin de case geselecteerd wordt, het inwinnen en verstrekken van gegevens, is de afdeling IVG, Inwinnen en Verstrekken Gegevens, de belangrijkste actor. Binnen deze afdeling zijn procesdeskundigen, procescoördinatoren, servicemanagers en functioneel beheerders de belangrijkste belanghebbenden bij de voorbereiding en uitvoering van het proces. Zij zullen vertegenwoordigd zijn in de klankbordgroep en een rol krijgen in de beoordeling van de methode. In de voortbrenging van de geautomatiseerde ondersteuning van processen binnen dit domein zijn procesontwerpers, architecten en applicatiebeheerders de belangrijkste betrokken actoren.

De klankbordgroep krijgt het ontwerp gepresenteerd in een plenaire bijeenkomst. Daarna vergelijkt de groep het nieuwe model met het oude en beoordeelt de toepasbaarheid. Hier worden gestructureerde en semigestructureerde interviews ingezet als methode om data te verzamelen en krijgt de klankbordgroep de ruimte om vragen te stellen tijdens de plenaire bijeenkomst. De interviewvragen richten zich op hoe begrippen uit de vraagstelling gemeten kunnen worden. In dit onderzoek gaat het om in hoeverre de beïnvloeding van datakwaliteit in het informatieverwerkende proces zichtbaar gemaakt kan worden om daarmee vanuit een procesperspectief de kwaliteit van data te verbeteren. Hiervoor kunnen de criteria gebruikt worden die in het literatuuronderzoek zijn verzameld en die in het praktijkonderzoek getoetst worden door de klankbordgroep. Het gaat hierbij om de eisen gericht op datakwaliteit in procesmodellen zoals die in paragraaf 2.3.5 zijn genoemd.

### 3.2.3. Stap 5, Evaluatie

In deze stap wordt beoordeeld of de nieuwe methode de organisatie daadwerkelijk handvatten geeft om te beoordelen of de beïnvloeding van datakwaliteit in het informatieverwerkende proces zichtbaar gemaakt kan worden om daarmee vanuit een procesperspectief de kwaliteit van data te verbeteren. Het is mogelijk dat in deze stap duidelijk wordt dat de ontwikkelde methode bijgesteld moet worden, dus dat vanaf stap 3 het plan opnieuw uitgevoerd moet worden. De aanpak is als volgt:

- De resultaten van de interviews die in stap 4 gehouden zijn, worden in eerste instantie uitgeschreven, getranscribeerd.
- Daarna worden ze samengevat (condenseren).
- In de analyse die daarop volgt wordt het ontwikkelde model getoetst aan de hand van vastgestelde criteria. De methode van analyse die hier het meest bij past is het vergelijken van patronen. Het kwantificeren van kwalitatieve data wordt door Saunders et al. (2012) genoemd aan aanvullende analysemogelijkheid en zal in dit onderzoek gebruikt gaan worden. Hiermee kan meetbaar gemaakt worden in hoeverre de nieuwe methode voldoet aan de criteria.
- Punten waar de nieuwe methode niet goed op scoort worden onderzocht. Is aanpassing van de methode mogelijk, dan wordt stap 3 en verder opnieuw voor dit deel uitgevoerd. Punten die niet of moeilijk aan te passen zijn omdat ze bijvoorbeeld eigen aan de gehanteerde taal zijn, worden gerapporteerd in het onderzoeksverslag.
- De evaluatie wordt teruggekoppeld aan de klankbordgroep en besproken. De resultaten van de terugkoppeling worden uitgeschreven en getranscribeerd. Ook dit kan aanleiding zijn om de methode bij te stellen en terug te gaan naar stap 3.
- In het kader van de beschikbare tijd van het onderzoek wordt een maximum aantal iteraties van twee gehanteerd. Punten die dan nog open staan worden gerapporteerd in het onderzoeksverslag.

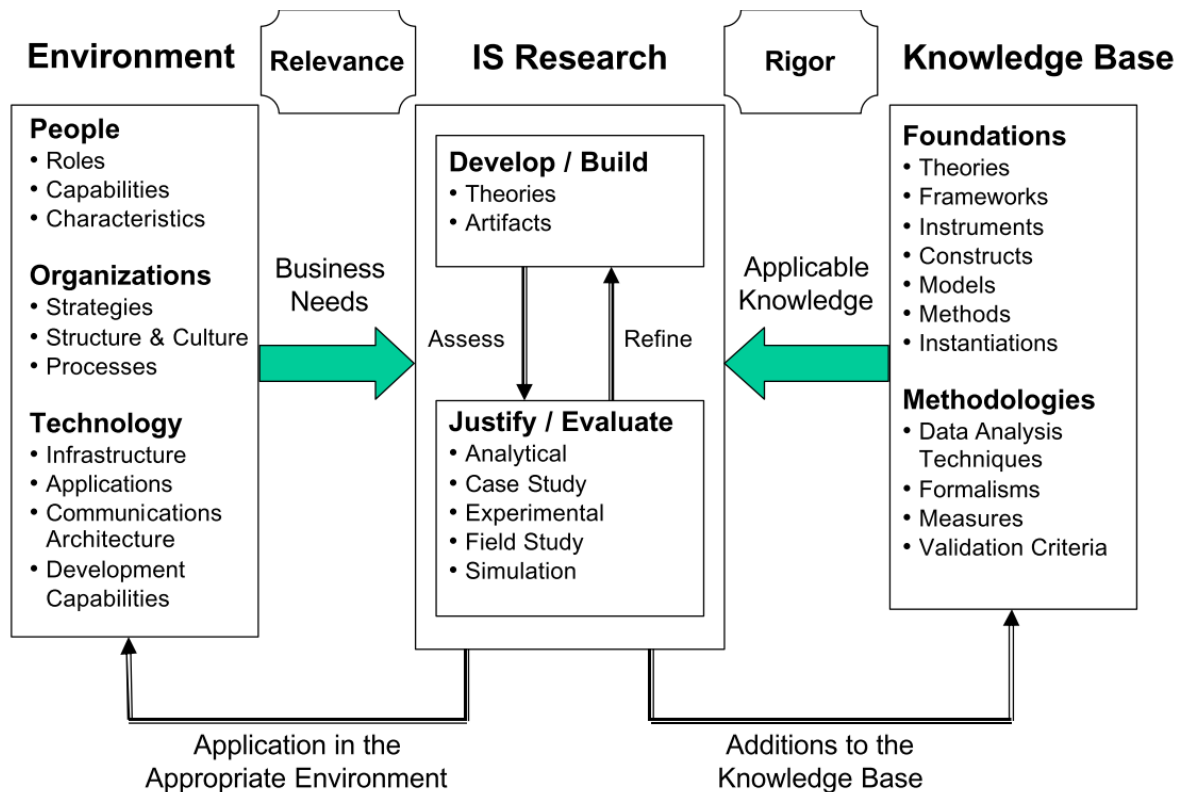
### 3.2.4. Stap 6, Communicatie

Deze stap bevat de uiteindelijke publicatie en presentatie van het onderzoek aan de OU en aan de onderzoeksorganisatie.

## 3.3. Methodologische issues

Zoals in de inleiding van dit hoofdstuk al genoemd stellen Peffers et al. (2007) dat veel onderzoek in ICT is gericht op het begrijpen en verklaren van fenomenen zoals dat in natuurwetenschappen

gebruikelijk is, maar dat die werkwijze minder geschikt is voor situaties waarin problemen opgelost moeten worden zoals ook in dit onderzoek het geval is waarin een nieuwe methode ontwikkeld moet worden. Daarmee worden de traditionele waarden als validiteit en betrouwbaarheid anders benaderd. Deze zullen in deze paragraaf dan ook anders ingevuld worden. Ook Hevner, March, Park, & Ram (2004) geven dat voor ICT onderzoek een combinatie van gedragswetenschappen, waar het vinden van waarheid centraal staat, en design science, waar gezocht wordt naar nut, geschikt is. De samenhang tussen deze twee benaderingen maken Hevner et al. (2004) duidelijk in hun 'Information Systems Research Framework' die in onderstaande figuur 7 is overgenomen. Hierin worden ook de twee belangrijkste methodologische issues in deze onderzoeksbenadering zichtbaar, 'rigor' en 'relevance'.



Figuur 7 Information Systems Research Framework van Hevner et al. (2004)

De linkerkolom Environment, de omgeving, geeft het probleemgebied weer. Vanuit mensen, organisaties en techniek ontstaan behoeften, business needs, in termen van doelen, taken, problemen en kansen. Deze business needs vormen het onderzoeksdomein en moeten een wetenschappelijke relevantie hebben. Hevner et al. (2004) werken deze relevantie uit in een van de zeven richtlijnen voor ICT onderzoek. In richtlijn 2, probleemrelevantie, wordt het doel genoemd van ICT onderzoek, het verkrijgen van kennis en inzicht waarmee de ontwikkeling en implementatie van op technologie gebaseerde oplossingen voor tot dan toe onopgeloste en belangrijke bedrijfsproblemen. Design research bereikt dit doel door het maken van innovatieve producten gericht op het veranderen van verschijnselen die optreden. In het literatuuronderzoek is aandacht besteed aan praktische en wetenschappelijke relevantie van dit onderzoek.

Rigor, uitgewerkt in richtlijn 5, wordt bereikt door effectief gebruik te maken van bestaande methoden en inzichten, de knowledge base (rechts in het framework) bij zowel de constructie van producten als het evalueren daarvan. Bij de constructie van het product wordt aansluiting gezocht bij methoden die al ontwikkeld zijn om datakwaliteit te verbeteren. In de evaluatie worden criteria

gebruikt die afkomstig zijn uit de literatuur en volgens een methode uit de literatuur (AHP) worden gewogen.

Aken, Chandrasekaran, & Halman (2016) noemen issues bij design research die in lijn liggen met de benadering van Hevner et al. (2004): hoe sterk is het bewijs dat het ontwerp dat gemaakt is daadwerkelijk het gewenste resultaat levert (pragmatische of interne validiteit) en wat is de praktische relevantie van het ontwerp. Het literatuuronderzoek gaat zoals genoemd in op zowel de wetenschappelijke als de praktische relevantie. De pragmatische of interne validiteit kan bereikt worden door het ontwerp uitgebreid te testen in het werkveld. De scope van dit onderzoek is beperkt tot het testen van één case in één organisatie, maar een aanbeveling kan zijn om het ontwerp op meerdere plaatsen binnen en buiten de onderzoeksorganisatie te testen. Aken et al. (2016) maken onderscheid tussen alfa testen, test door de onderzoekers zelf, en bèta testen, testen door stakeholders, in dit onderzoek de klankbordgroep. Het testen kan aangevuld worden door peer reviews of gesprekken met experts. In dit onderzoek worden experts van de centrale afdeling methoden en technieken betrokken in het onderzoek.

Dit onderzoek raakt geen ethische kwesties. Er wordt een probleem onderzocht waarbij morele waarden geen rol spelen.

## 4. Resultaten

In dit hoofdstuk worden de resultaten van het onderzoek weergegeven. Hierbij worden de activiteiten e en f van stap 3 uit het plan van aanpak uitgewerkt.

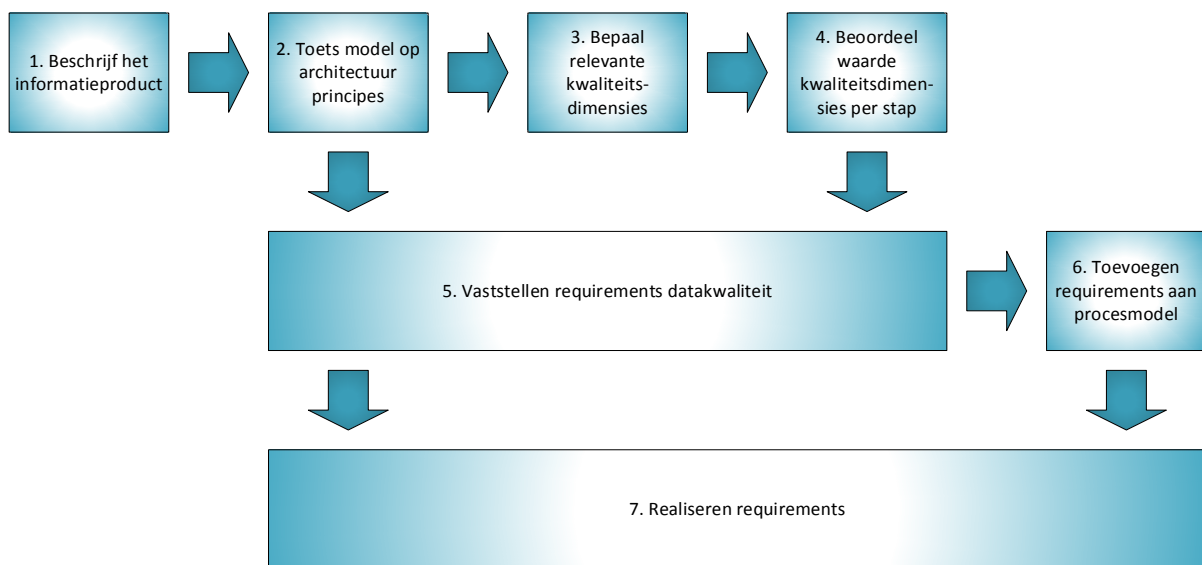
### 4.1. Selecteren modelleertaal

De keuze voor een modelleertaal als basis voor de methode is bepaald door de case in de onderzoeksorganisatie. Daar wordt BPMN gebruikt, een standaard taal voor procesmodelleren die wereldwijd gebruikt wordt en beheerd wordt door de Object Management Group. Door deze taal te gebruiken in de methode voldoet deze aan een van de uitgangspunten die in het literatuuronderzoek zijn vastgesteld (paragraaf 2.3.5), namelijk dat de modellen aan moeten sluiten bij standaard BPM methoden.

### 4.2. Ontwerp van de methode DQPM

#### 4.2.1. Overview van de methode

Op basis van de literatuur en het bestuderen van de case is onderstaande methode ontwikkeld. De naam van de methode is DQPM, Data Quality Process Modelling.



Figuur 8 Overview DQPM

#### Toelichting overview DQPM

Het theoretisch kader van de methode wordt in paragraaf 4.2.2 beschreven en waar nodig in de uitwerking per stap. De stappen worden in paragraaf 4.2.4 verder uitgewerkt voorafgegaan door een aantal algemeen geldende uitgangspunten (paragraaf 4.2.3). Hieronder een korte beschrijving en een toelichting op de overview.

1. De eerste stap beschrijft het informatieproduct dat door het proces voortgebracht wordt, de eisen die afnemers er aan stellen en de plek in het proces waar het product gecreëerd wordt. Dit gebeurt op het niveau van werkproces.
2. In de tweede stap wordt gekeken naar het volledige bedrijfsproces waar het werkproces onderdeel van is en wordt beoordeeld of het model voldoet aan belangrijke

architectuurprincipes zoals 'quality at source', het vaststellen van kwaliteit van gegevens aan het begin van het proces. Afhankelijk van de uitkomst levert dit requirements op voor aanpassing van het proces.

3. De volgende stap, stap 3, wordt weer uitgevoerd op het niveau van werkproces en omvat het vaststellen van relevante kwaliteitsdimensies. Deze zijn afhankelijk van het doel van het proces en de eisen van de relevante stakeholders.
4. Van die dimensies wordt in stap 4 beoordeeld of deze in voldoende mate in het proces zijn geborgd.
5. Tekortkomingen worden bij stap 5 vastgelegd in requirements die afzonderlijk beschreven en gemodelleerd worden.
6. In stap 6 worden ze ook gerelateerd aan het procesmodel.
7. Het realiseren van de requirements (stap 7) vindt plaats volgens de geldende procedures binnen de organisatie.

#### 4.2.2. Onderbouwing van de methode

Uitgangspunt voor de methode is de procesoriëntatie zoals die onder meer in de Code voor Informatiekwaliteit (Vieveen et al., 2016) terug te vinden is. Informatieverwerking wordt daarin als een proces gezien, het denkraam voor het managen van informatiekwaliteit is gebaseerd op de benadering die in de productie-industrie wordt gebruikt. Een proces wordt door Obers & Achterberg (2014) gedefinieerd als "een samenhangend geheel van activiteiten, mensen en middelen, waarmee een of meer producten of diensten worden voortgebracht". Het product van een informatieverwerkend proces wordt een 'informatieproduct' genoemd, "een set digitale gegevens in een vast te pakken vorm" (Besselink, 2010).

De basis voor DQPM ligt in de fasering die in het literatuuronderzoek beschreven is. De inventarisatie van Batini et al. (2009) geeft aan dat de meeste methoden om datakwaliteit te verbeteren in een procesaanpak drie fases bevatten:

a. **Beschrijvend.**

Stap 1 en 3 van DQPM passen in deze fase. Hierin wordt gekeken naar het product dat voortgebracht wordt door het proces en de kenmerken daarvan en er wordt gekeken naar welke kwaliteitsdimensies voor het te beoordelen proces relevant zijn.

b. **Assessment.**

Stap 2 en 4 van DQPM passen in deze fase. Stap 2 kijkt meer naar de buitenkant van het proces door het tegen een aantal architectuurprincipes aan te houden. Daarmee wordt beoordeeld of het proces intrinsiek al een aantal kenmerken heeft waardoor de kwaliteit van data positief of negatief beïnvloed wordt. Als controles bijvoorbeeld pas aan het eind van het proces zijn opgenomen, brengt dit risico's met zich mee voor de kwaliteit van de gegevens (Shankaranarayanan et al., 2000).

Stap 4 kijkt meer van binnenuit naar de kwaliteit door per stap in het proces te beoordelen hoe de kwaliteitsdimensies afgedekt worden.

c. **Verbetering.**

Stap 5, 6 en 7 geven invulling aan verbetering. Eerst worden requirements vastgesteld, dan worden die toegevoegd aan het procesmodel zodat de soll situatie beschreven is en dan worden ze gerealiseerd. Het formuleren van requirements om de verbetering in te vullen is gebaseerd op het artikel van Rodríguez, Caro, Capiello, & Caballero (2012).

### 4.2.3. Uitgangspunten voor gebruik van de methode

Om de methode succesvol toe te kunnen passen in een organisatie, wordt er vanuit gegaan dat aan onderstaande uitgangspunten voldaan wordt.

1. De methode gebruikt Business Process Model and Notation (BPMN) als basis. Aanname is dat degenen die betrokken zijn bij de uitvoering van de methode, modellen die in deze taal gemaakt zijn kunnen lezen.
2. Er is een procesmodel beschikbaar en andere relevante documentatie van het proces zoals ontwerpdocumentatie (functioneel en technisch), een conceptueel gegevensmodel waarin de betekenis van de in het proces gebruikte gegevens beschreven wordt. Ook informatie over de uitvoering van het proces is zinvol: aantal storingsmeldingen, gemelde incidenten, doorlooptijden etc.
3. Het procesmodel is een afspiegeling van de werkelijkheid. Voordat het model gebruikt wordt in de methode moet het getoetst worden met deskundigen uit het proces.
4. Deze methode gaat er vanuit dat er een proces beschikbaar is om beoordeeld te worden. Welk proces beoordeeld gaat worden en hoe dat ingebed is in kwaliteitsprocedures, is aan de organisatie die de methode toepast. Daarbij kunnen wel de criteria gehanteerd worden die ook bij het selecteren van de case van dit onderzoek gebruikt zijn:
  - a. het proces moet meerdere gegevensbewerkingen bevatten;
  - b. het proces moet in ontwikkeling zijn;
  - c. er moeten verschillende stakeholders betrokken zijn bij dit proces;
  - d. het proces heeft bij voorkeur problemen met datakwaliteit.

Aandachtspunt hierbij kan nog zijn welke gevalscategorieën die door het werkproces behandeld worden geselecteerd worden. Generieke processen kunnen meerdere typen gevallen behandelen waar voor de methode mogelijk een keuze uit gemaakt moet worden. Ook die keuze is aan de organisatie en valt buiten scope van de methode DQPM.

5. Met uitzondering van stap 2 is het beschouwingsgebied van deze methode het niveau van werkproces en de raakvlakken met andere processen. Stap 2 kijkt ook naar de bovenliggende niveaus van bedrijfsproces en ketenproces. Deze gebruikte procesdecompositie en bijbehorende definities is terug te vinden in diverse interne voorschriften van de onderzoeksorganisatie, de NORA (Nederlandse Overheid Referentie Architectuur) en in Grip op processen in organisaties van Obers & Achterberg (2014). In het kort:
  - a. Ketenproces: een samenhangend geheel van bedrijfsprocessen waarin meerdere organisaties een rol spelen om samen een product of dienst voort te brengen.
  - b. Bedrijfsproces: een geordende reeks werkprocessen binnen één organisatie die een product of dienst leveren.
  - c. Werkproces: een samenhangende reeks processtappen die de behandeling van een gevalscategorie beschrijft en die een bijdrage aan een product of dienst levert.
  - d. Processtap: een eenheid van tijd, actor en besturing.
  - e. Activiteit: niveau van handeling binnen een processtap.
6. De stakeholders van het proces zijn bekend en beschikbaar.



## 4.2.4. Toelichting per stap

### *Stap 1 – Beschrijf het informatieproduct*

#### **Actie**

Beschrijf het informatieproduct dat door het werkproces geleverd wordt. De beschrijving bevat de volgende elementen:

- Uit welke onderdelen bestaat het product. Als er een conceptueel gegevensmodel (CGM) beschikbaar is, is het praktisch om aan te sluiten bij de concepten die in dat model beschreven zijn.
- Wat is de betekenis van de gegevens die in het informatieproduct zitten. Ook hier kan de bron een CGM zijn.
- Wie zijn de afnemers van het product?
- Welke eisen stellen afnemers aan het product?
- Markeer waar in het proces het eindproduct gereed is. Maak de betreffende processtap rood.

In een proces worden op verschillende plaatsen gegevens gebruikt, verwerkt, gemuteerd of verzonden naar andere processen. Het gaat om de verschillende concepten in BPMN. Om te bepalen waar een informatieproduct wordt gemaakt, zijn alleen taken waarin een zogenaamde data store of een data object gebruikt worden van belang. Een overzicht van BPMN concepten waarin gegevensgebruik voorkomt is opgenomen in bijlage 6.

Aandachtspunt: uit de toepassing van de methode voor het werkproces ‘Controleren ingewonnen gegevens’ en uit een quickscan van een vijftal andere modellen uit de onderzoeksorganisatie blijkt dat er ook processtappen of activiteiten zijn waarin data-items niet gemodelleerd zijn, maar waar wel degelijk data gebruikt of bewerkt wordt. Dan moet uit de naamgeving blijken of data gebruikt of bewerkt wordt en of er dus een informatieproduct gecreëerd wordt.

#### **Onderbouwing**

De methode richt zich op het zichtbaar maken van de beïnvloeding van datakwaliteit van een informatieproduct. Aan de hand van het metamodel van BPMN (Object Management Group (OMG), 2011) wordt onderzocht waar het informatieproduct in het proces gecreëerd wordt, hoe het informatieproduct eruit ziet en welke eisen afnemers stellen aan het product.

#### **Resultaat**

Een beschrijving van het informatieproduct dat door het werkproces geleverd wordt en een indicatie waar het product in het proces gecreëerd wordt.

### *Stap 2 – Toets model op architectuurprincipes*

#### **Actie**

In de literatuur en in de praktijk van de Belastingdienst zijn verschillende principes te vinden die de kwaliteit van gegevens bevorderen. In deze stap wordt het model getoetst op het naleven van deze principes. Om deze principes te toetsen is het nodig om niet alleen naar het niveau van werkproces of processtap te kijken, maar ook naar het niveau van bedrijfsproces en ketenproces. Zie de uitgangspunten voor de definities en samenhang tussen deze niveaus. De principes die vanuit DQPM

aangereikt worden zijn hieronder genoemd. Daarnaast kan het proces getoetst worden op specifieke binnen de organisatie of het project geldende architectuurprincipes die betrekking hebben op datakwaliteit.

- Quality at source (Shankaranarayanan et al., 2000). Dit betekent, stel de kwaliteit van de gegevens zoveel mogelijk in het begin van het proces vast, dan is eventueel herstel eenvoudiger en goedkoper.
- Modelleer de gehele keten (Wokke, 2018). Door de gehele keten te modelleren wordt inzicht verkregen in hoe de kwaliteit bepaald of beïnvloed wordt in andere delen van de keten. Daarmee wordt data lineage mogelijk gemaakt, het laten zien wat er met data gebeurt door de tijd heen in processen en systemen.
- Breng de gebruiker naar de data, niet de data naar de gebruiker (Otten & Karels, 2018). Dit voorkomt onnodig 'gesleep' van data waar risico's aan verbonden zijn.
- Hieraan gerelateerd: enkelvoudig beheer, meervoudig gebruik (genoemd tijdens klankbordsessie).
- Het besturingsproces is ingericht (Vieveen et al., 2016). Dit zorgt voor het kunnen beheersen van het proces.
- Het informatieverwerkingsproces is volledig weergegeven, dat wil zeggen dat het invoerproces, de verwerking van input tot output, het bewaren, archiveren en vernietigen van datasets en het leveren van output aan de afnemer is beschreven in het proces (Vieveen et al., 2016).
- Het informatieverwerkingsproces is zo efficiënt mogelijk ingericht, dat wil zeggen dat IT optimaal is ingezet en dat er gebruik gemaakt wordt van standaards en generieke diensten (Vieveen et al., 2016).
- Er zijn processen geïmplementeerd voor het corrigeren van fouten in de throughput en output (Vieveen et al., 2016).
- De raakvlakken van het onderzochte werkproces met andere processen zijn zichtbaar gemaakt in het model. Van de afnemende werkprocessen zijn de kwaliteitseisen beschreven die gesteld worden aan het informatieproduct (genoemd tijdens klankbordsessie).

Stel voor de niet-nageleefde principes requirements op hoe wel aan het principe voldaan kan worden.

### **Onderbouwing**

Zie bronverwijzing bij de principes.

### **Resultaat**

Een overzicht van principes die wel of niet nageleefd zijn in het model en requirements om wel aan de principes te kunnen voldoen.

### *Stap 3 – Bepaal relevante kwaliteitsdimensies*

#### **Actie**

Geef aan welke kwaliteitsdimensies voor het proces en vervolgens per processtap relevant zijn. De dimensies voor het proces kunnen met de procesverantwoordelijke en andere stakeholders vastgesteld worden. Neem van elke dimensie de definitie op. Soms kan de dimensie per processtap

of voor het specifieke geval dat behandeld wordt in het werkproces een andere betekenis hebben. Neem dan niet alleen de definitie op zoals die geldt voor het werkproces, maar ook voor de processtap of voor het geval. Voor de kwaliteitsdimensies per processtap kan de volgende tabel gebruikt worden. Hierbij is de tabel uit bijlage 1 als basis gebruikt.

Tabel 3 Inventarisatie kwaliteitsdimensies per soort processtap

| Soort                 | Kwaliteitsdimensie              | Lezen/<br>catching<br>throwing | Controleren | Muteren | Verwijderen |
|-----------------------|---------------------------------|--------------------------------|-------------|---------|-------------|
| Intrinsieke kwaliteit | <b>Juistheid</b>                | X                              | X           | X       | X           |
|                       | <b>Geloofwaardigheid</b>        | X                              | X           | X       | X           |
|                       | <b>Consistentie</b>             | X                              | X           | X       | X           |
| Context               | <b>Relevantie</b>               | X                              | X           | X       | X           |
|                       | <b>Tijdigheid</b>               | X                              | X           | X       | X           |
|                       | <b>Volledigheid</b>             | X                              | X           | X       | X           |
| Representatie         | <b>Interpreteerbaarheid</b>     | X                              | X           |         |             |
|                       | <b>Semantische consistentie</b> | X                              | X           |         |             |
|                       | <b>Begrijpelijkheid</b>         | X                              | X           |         |             |
| Toegankelijkheid      | <b>Toegankelijkheid</b>         | X                              | X           |         |             |
|                       | <b>Beveiliging</b>              | X                              | X           |         |             |

#### Toelichting

- Kolom Lezen/catching/throwing: catching en throwing worden beide gezien als een vorm van lezen. Een intermediate throwing event doet hetzelfde als een send task (Object Management Group (OMG), 2011) en die task verzendt alleen maar de gegevens, er vindt geen bewerking meer plaats, dat is mogelijk in de stap daarvoor gedaan.
- Kolom Controleren: dit is een bijzondere vorm van lezen, te weten lezen met een bepaald doel; een oordeel geven en vastleggen over het gegeven, bijvoorbeeld ‘goed’, ‘fout’ of ‘gedeeltelijk bruikbaar’. Een relatie in de kolom is afhankelijk van het aspect waarop gecontroleerd wordt. Als bijvoorbeeld alleen de juistheid van een bankrekeningnummer wordt gecontroleerd door middel van de elfproef, dan zegt dat niets over de volledigheid. Daarom staan de waardes in deze kolom grijs weergegeven.

#### Onderbouwing

De invulling van de tabel is gedaan aan de hand van de definities van de dimensies zoals die in de bijlage 1 van het onderzoek zijn opgenomen.

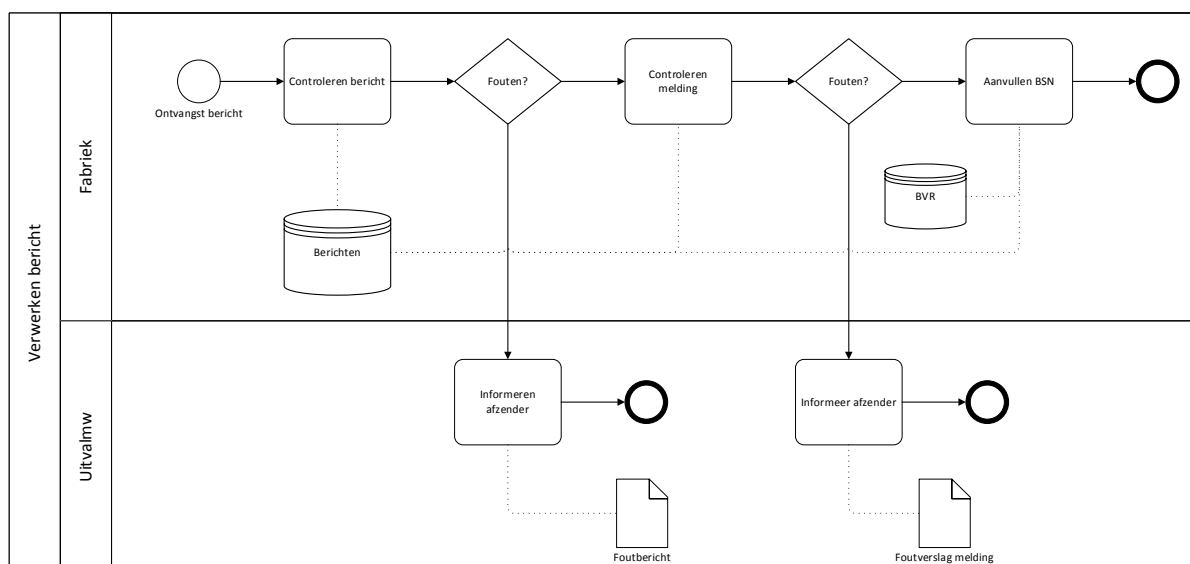
#### Resultaat

Een overzicht van welke stappen een relatie hebben met welke kwaliteitsdimensie. Het overzicht wordt in een tabel gezet worden met in de kolommen de kwaliteitsdimensie en op de rijen de stappen of events in het proces. Dat ziet er als volgt uit:

Tabel 4 Resultaat inventarisatie kwaliteitsdimensies per processtap

| Werkproces ABC    | Processtap 1 | Processtap 2 | Intermediate event X | Processtap 3 | Etc. etc. |
|-------------------|--------------|--------------|----------------------|--------------|-----------|
| Juistheid         | X            |              | X                    |              | ...       |
| Geloofwaardigheid | X            |              |                      |              | ...       |
| Consistentie      | X            |              | X                    |              | ...       |
| Relevantie        | X            | X            | X                    | X            | ...       |
| Etc. etc.         | ...          | ...          | ...                  | ...          |           |

De definities van de dimensies worden na de tabel opgenomen. Geef onder de tabel ook aan welke dimensies niet gebruikt zijn in dit proces. Daarnaast wordt een grafische weergave opgesteld die wordt toegevoegd onderaan het model waarin de dimensies staan. Dat ziet er als volgt uit (NB, ander voorbeeld dan de tabel hierboven):



| Kwaliteitsdimensie/Proces | CB | IA | CM | IA |
|---------------------------|----|----|----|----|
| Volledigheid              | X  |    |    |    |
| Juistheid                 | X  |    |    | X  |
| Geloofwaardigheid         |    | X  |    | X  |
| Betrouwbaarheid           |    |    | X  |    |
| Etc. etc.                 |    |    |    |    |

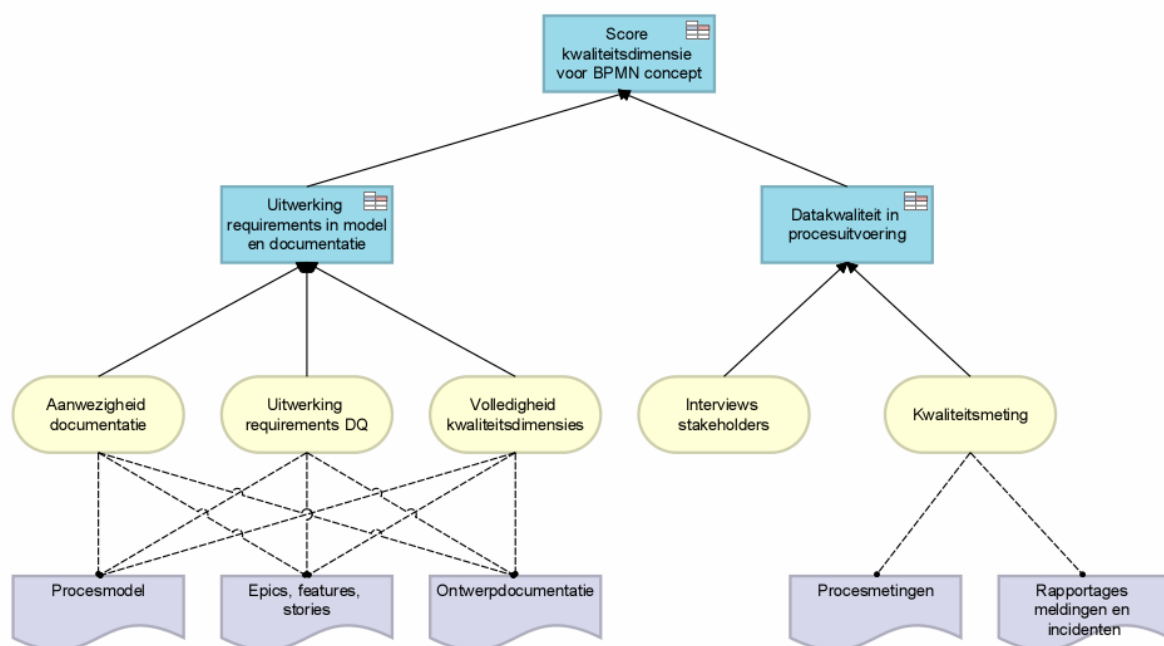
Figuur 9 Resultaat inventarisatie kwaliteitsdimensie voor procesmodel

Het bovenste gedeelte is het standaard BPMN procesmodel. Het onderste gedeelte, de 'notenbalk', bevat de voor dat proces relevante kwaliteitsdimensies en of ze een relatie hebben met de processtap of het event erboven. Op de eerste regel van de 'notenbalk' is een verwijzing opgenomen naar de processtap of ander onderdeel van het proces waar de relatie betrekking op heeft. Dit is gedaan door de eerste letters van de processtap op te nemen. Bijvoorbeeld CB staat voor de processtap 'Controleren bericht'.

#### Stap 4 – Beoordeel waarde kwaliteitsdimensies per stap – de baselinesituatie

##### Actie

Stel aan de hand van de gemodelleerde activiteiten in de stap, de beschrijving van het model en andere informatiebronnen vast in welke mate requirements met betrekking tot kwaliteit van gegevens duidelijk zijn en uitgewerkt zijn in het proces. Daarvoor kan onderstaand beslisdiagram met bijbehorende beslistabellen gebruikt worden. Deze zijn gemodelleerd in Decision Model and Notation (DMN), een open standaard zoals BPMN om beslissingen te modelleren en beheerd door de Object Management Group die ook BPMN beheert. Naast het procesmodel met de bijbehorende beschrijving kan ook ontwerpdocumentatie zoals een overzicht van de gerealiseerde features en userstories, een functioneel ontwerp of een UML diagram geraadpleegd worden. Verder kan verzamelde informatie over de procesuitvoering gebruikt worden zoals het aantal incidenten, storingen, metingen en interviews met uitvoerenden. De score wordt bepaald met de stakeholders die in de voorbereidende stap verzameld zijn. In eerste instantie wordt een beeld per processtap verkregen. De score wordt opgenomen in de tabel of het grafische overzicht uit de vorige stap en de daarbij geselecteerde kwaliteitsdimensies. Daarbij gelden de volgende richtlijnen om te scoren. De aanname bij deze richtlijnen is voor bestaande processen daarbij wel dat de bouw conform ontwerp is.



Figuur 10 Decision requirements diagram score kwaliteitsdimensies voor procesmodel

##### Toelichting

De beslissing om een score voor een kwaliteitsdimensie toe te kennen aan een BPMN concept (veelal een processtap, maar dat kan bijvoorbeeld ook een intermediate event zijn) is gebaseerd op twee andere beslissingen: de mate waarin requirements uitgewerkt zijn in het model en de documentatie en de mate waarin datakwaliteit in uitvoering van het proces vastgesteld wordt.

De eerste beslissing, de mate waarin requirements uitgewerkt zijn in het procesmodel en de documentatie heeft drie inputgegevens: de aanwezigheid van documentatie, de uitwerking van requirements op het gebied van datakwaliteit en de volledigheid van de kwaliteitsdimensies die al

dan niet uitgewerkt zijn. Die input is onder meer te vinden in procesmodellen, epics, features, stories en ontwerpdocumentatie. De tabel om de beslissing af te leiden is opgenomen in bijlage 4.

Afhankelijk van of de documentatie aanwezig is of niet, de requirements uitgewerkt, afleidbaar of niet uitgewerkt en niet afleidbaar zijn en de kwaliteitsdimensies volledig, gedeeltelijk of niet zijn uitgewerkt, is het resultaat van de beslissing of requirements uitgewerkt zijn in het model en documentatie goed, redelijk of slecht.

De tweede deelbeslissing betreft de mate waarin datakwaliteit in uitvoering van het proces vastgesteld wordt. De bijbehorende beslistabel is opgenomen in bijlage 4.

Deze beslissing is gebaseerd op de informatie die verkregen wordt uit interviews met uitvoerenden en beheerders van het proces en metingen die in het proces gedaan worden. Metingen kunnen direct op het proces uitgevoerd worden, zoals metingen van aantallen op plekken in het proces, het uitlezen van loggegevens, etc., maar ook meer indirect door te kijken naar gemelde incidenten en storingen. De uitkomst van deze deelbeslissing is goed, redelijk of slecht.

De uitkomsten van de twee deelbeslissingen leiden samen tot een oordeel over de score van kwaliteitsdimensies voor een BPMN concept. De bijbehorende beslistabel is opgenomen in bijlage 4. Hierbij weegt de datakwaliteit in uitvoering iets zwaarder dan de uitwerking van requirements in het model en de documentatie, omdat documentatie achter kan lopen bij de werkelijkheid. Een aanname (zie uitgangspunten) is dat de documentatie en het model actueel is, maar in de praktijk is dit niet altijd te garanderen.

Als de score voor een processtap is vastgesteld met deze laatste tabel, moet nog gekeken worden welke dimensies goed scoren en welke niet. Daarvoor kan onderstaande afleiding gebruikt worden:

*Tabel 5 Beslissing waarde per dimensie voor BPMN concept*

| Score BPMN concept                                    | Waarde per dimensie voor BPMN concept   |
|---|---|
| Goed  | Alle dimensies scoren goed  |
| Redelijk, uitwerking kwaliteitsdimensies gedeeltelijk | Dimensies die uitgewerkt zijn scoren goed<br>Dimensies die niet uitgewerkt zijn scoren slecht     |
| Redelijk, geen uitwerking kwaliteitsdimensies         | Alle dimensies scoren slecht.   |
| Slecht, uitwerking kwaliteitsdimensies gedeeltelijk   | Dimensies die uitgewerkt zijn scoren redelijk<br>Dimensies die niet uitgewerkt zijn scoren slecht |
| Slecht, geen uitwerking kwaliteitsdimensies           | Alle dimensies scoren slecht.   |

## Onderbouwing

Rodríguez et al. (2012) gaan in hun BPMN extensie uit van requirements als sleutel om datakwaliteit te borgen in het proces. Daarmee wordt het probleem opgelost dat voor nieuwe processen die nog niet geïmplementeerd zijn de kwaliteit van data niet gemeten kan worden. In de praktijk van de Belastingdienst blijkt ook dat het meten van kwaliteit een complexe aangelegenheid is waar niet op korte termijn inzicht in de waarde van de dimensies verkregen kan worden. Door na te gaan hoe kwaliteit geborgd is in het proces wordt dit probleem tegengegaan.

## Resultaat

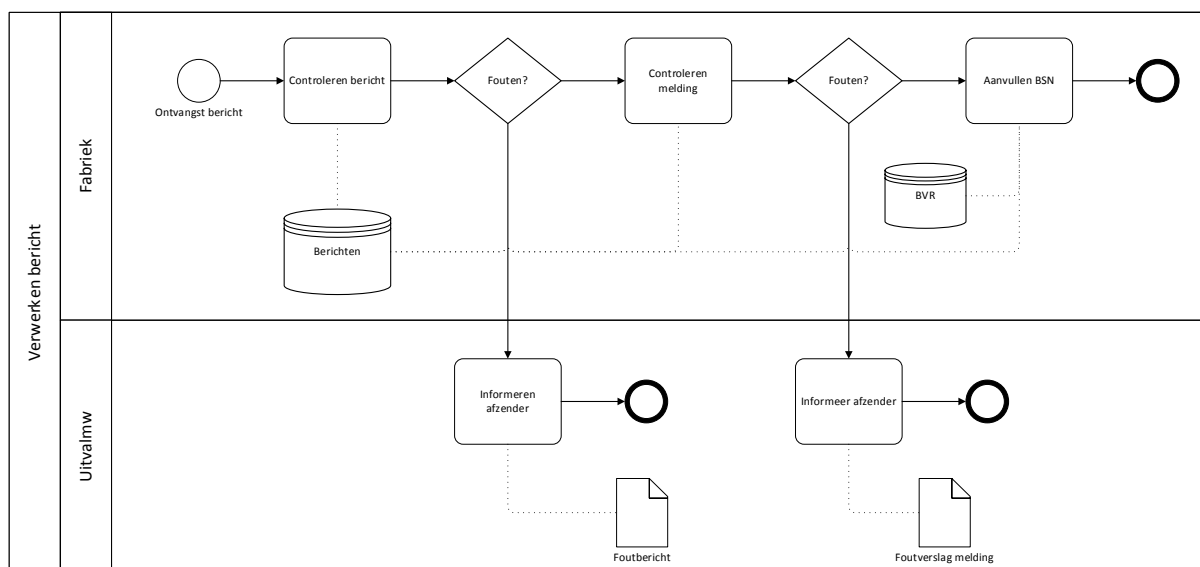
Het resultaat is dat per processtap een oordeel gegeven is in hoeverre requirements uitgewerkt en toegepast zijn met betrekking tot datakwaliteit. Dit wordt analoog aan de vorige stap op twee manieren gevisualiseerd, in een tabel en als toevoeging op het procesmodel.

In de tabel worden de smileys vervangen door +, - en +/- . Dat ziet er dan als volgt uit:

Tabel 6 Score kwaliteitsdimensies voor procesmodel

| Werkproces ABC       | Juistheid | Geloofwaardig-<br>heid | Consistentie | Relevantie | Etc. etc. |
|----------------------|-----------|------------------------|--------------|------------|-----------|
| Processtap 1         | +         | -                      | +/-          | +          | ...       |
| Processtap 2         |           |                        |              | +          | ...       |
| Intermediate event X | +/-       |                        | -            |            | ...       |
| Processtap 3         |           |                        |              | +/-        | ...       |
| Etc. etc.            | ...       | ...                    | ...          | ...        | ...       |

Op de grafische manier weergegeven als toevoeging op het procesmodel ziet het er als volgt uit:



| Kwaliteitsdimensie/Proces | CB | CM | AB |
|---------------------------|----|----|----|
| Volledigheid              | 😊  | 😊  |    |
| Juistheid                 | 😊  | 😊  |    |
| Geloofwaardigheid         |    |    |    |
| Betrouwbaarheid           |    |    | 😐  |
| Etc. etc.                 |    |    |    |

Figuur 11 Score kwaliteitsdimensies voor procesmodel

Het bovenste gedeelte is het standaard BPMN procesmodel. Het onderste gedeelte, de 'notenbalk', bevat de voor dat proces relevante kwaliteitsdimensies met een score voor de processtap of het event erboven. Op de eerste regel van de 'notenbalk' is een verwijzing opgenomen naar de processtap of ander onderdeel van het proces waar de score betrekking op heeft. Dit is gedaan door de eerste letters van de processtap op te nemen. Bijvoorbeeld CB staat voor de processtap

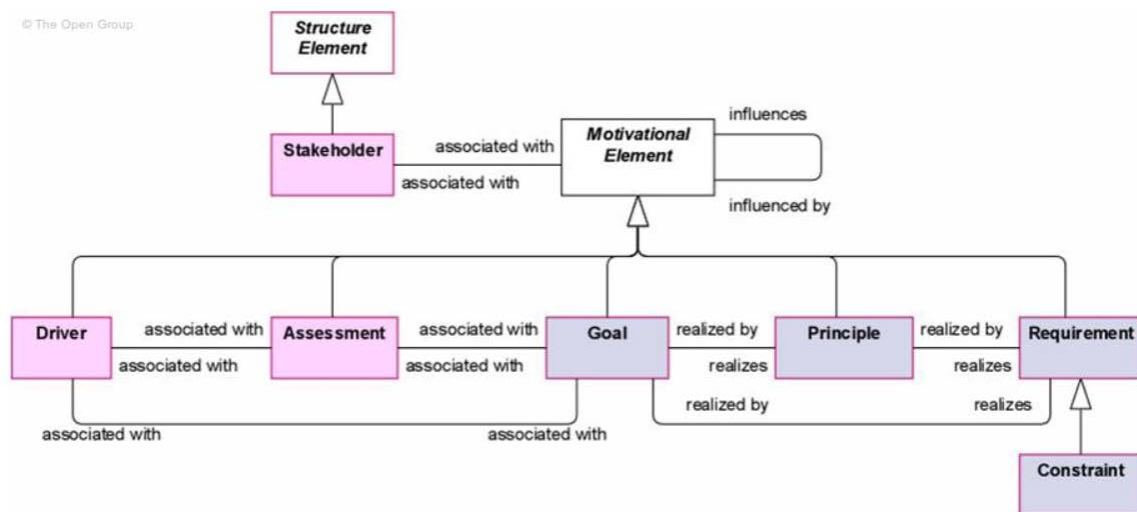
‘Controleren bericht’.

#### Stap 5 – Vaststellen requirements datakwaliteit

##### Actie

De vorige stap laat zien waar de datakwaliteit in het proces verbeterd kan worden door aanvullende eisen te stellen in de vorm van requirements. Dit kan als requirement vastgelegd worden, maar ook geformuleerd worden als user story. Belangrijk is dat het requirement gerealiseerd kan worden in het proces en in de IT ondersteuning.

Naast requirements kunnen ook principles of uitgangspunten geformuleerd worden zoals die in de tweede stap geformuleerd zijn, bijvoorbeeld ‘quality at source’ (Shankaranarayanan et al., 2000). In deze methode wordt de motivatie-extensie van Archimate gebruikt om requirements vast te leggen. Figuur 12 laat de concepten uit het metamodel zien. De gebruikte kleuren zijn van de opstellers, de betekenis daarvan is voor dit onderzoek niet van belang. Daaronder de toelichting op hoe deze gevuld moeten worden. Voor deze methode zijn niet alle concepten nodig.



Figuur 12 Metamodel motivatie-extensie (Object Management Group (OMG), 2013)

Voor deze methode zijn relevant:

- Stakeholder: de belanghebbende(n) van dit proces, bijvoorbeeld de proceseigenaar en de afnemer van het informatieproduct.
- Driver: gebruik hier de kwaliteitsdimensies.
- Assessment: beschrijf hier de tekortkomingen die uit de vorige stap naar voren zijn gekomen.
- Goal: belangrijkste doelen om tekortkomingen op te heffen.
- Requirement: wat is nodig om datakwaliteit te borgen of te verbeteren. Van een requirement wordt altijd vastgelegd wat de gewenste functionaliteit is en de rationale. Op deze plek kan ook een user story vastgelegd worden.
- Optioneel: principes en constraints

##### Onderbouwing

De reden om de motivatie-extensie te gebruiken is dat requirements niet alleen beschreven worden, maar dat de samenhang met knelpunten en stakeholders ook gevisualiseerd wordt en dat daarbij



een wereldwijd erkende standaard gebruikt wordt, Archimate. Verder is het met de ontwerptooling die de Belastingdienst gebruikt mogelijk om requirements uit de motivatie-extensie te koppelen aan concepten uit een BPMN procesmodel.

## **Resultaat**

Een verzameling requirements, beschreven en gevisualiseerd in de motivatie-extensie van Archimate, om de kwaliteit van data te borgen of te verbeteren.

### *Stap 6 – Toevoegen requirements aan procesmodel – de targetsituatie*

#### **Actie**

Deze stap kan afhankelijk van de tool waarmee gemodelleerd wordt anders ingevuld worden. In dit onderzoek wordt gebruik gemaakt van de tool die bij de onderzoeksorganisatie voorgeschreven is, BiZZdesign Enterprise Studio. In deze tool is het mogelijk om door middel van een modeloverstijgende relatie requirements uit de vorige stap, vastgelegd in de motivatie-extensie van Archimate, te koppelen aan processtappen, messages en events in BPMN en daar vervolgens over te rapporteren door middel van een tabel waarin de proceselementen op de rijen staan en de motivatie-elementen in de kolommen. Voor de IV voortbrenging is hiermee duidelijk welke requirements waar in het proces gesteld worden. N.B. Een requirement kan er ook toe leiden dat er extra stappen aan het proces toegevoegd moeten worden, bijvoorbeeld een controlestap.

#### **Onderbouwing**

In requirementsmanagement is het kunnen volgen van de lifecycle van requirements (traceability) belangrijk (Zou, Settimi, & Cleland-Huang, 2010). Door requirements te koppelen aan plekken in het proces wordt dit ondersteund. Verder wordt het effect van de getroffen maatregelen zichtbaar gemaakt, de targetsituatie.

## **Resultaat**

Een overzicht van waarin het proces welke requirements gesteld worden om de kwaliteit van data te verbeteren.

### *Stap 7 – Realiseren requirements*

#### **Actie**

In deze stap worden de requirements al dan niet gedeeltelijk gerealiseerd en wordt het procesmodel bijgewerkt als gevolg van het realiseren van de requirements. De stap bestaat uit een aantal deelactiviteiten.

- Selecteer welke requirements gerealiseerd gaan worden. Hoe die selectie plaatsvindt, wordt niet door deze methode voorgeschreven. De keuze kan beïnvloed worden door risico, budget, beschikbare capaciteit etc. Deze factoren zijn organisatiespecifiek en de selectie moet binnen de kaders en bestaande procedures van de organisatie uitgevoerd worden.
- Geef van niet gerealiseerde requirements het risico aan. Geef deze in de notenbalk een afwijkende witte kleur zodat zichtbaar is dat er een risico gelopen wordt. Geef aan waarom het requirement niet gerealiseerd wordt. Bijvoorbeeld omdat de organisatie het risico accepteert of omdat het om een known error gaat.

**Onderbouwing**

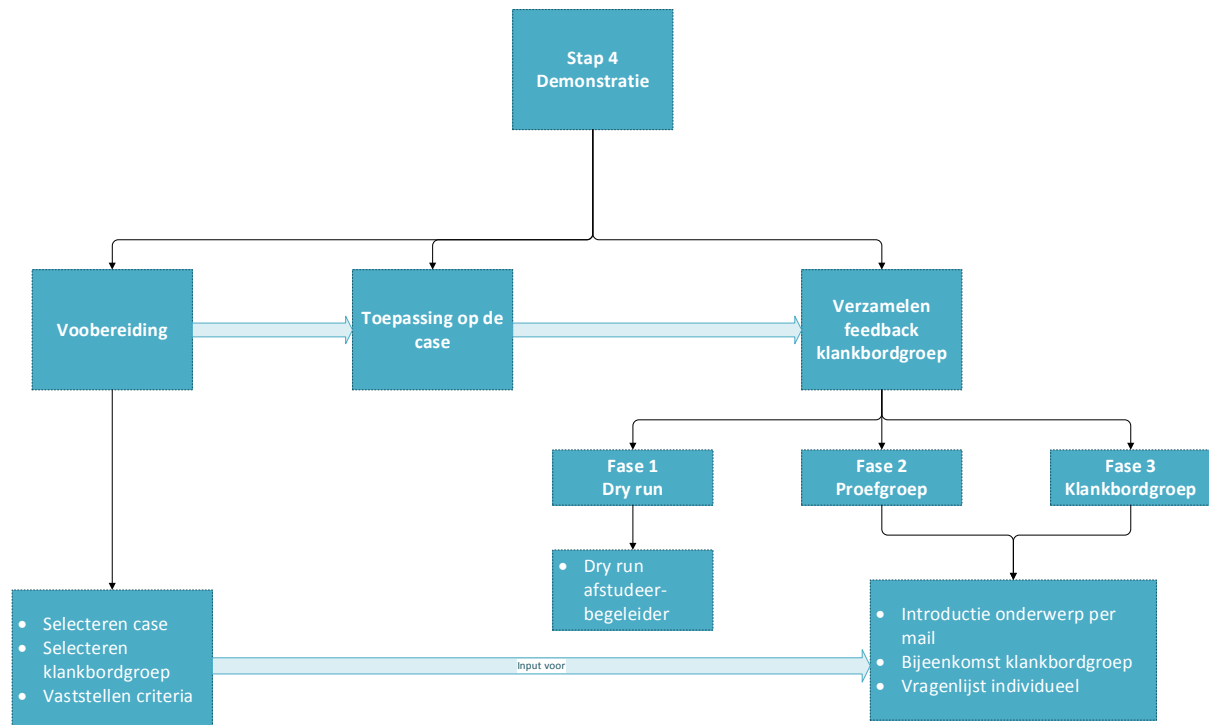
N.v.t.

**Resultaat**

Het resultaat is een proces dat beter in staat is om de kwaliteit van data te borgen.

## 5. Demonstratie: toepassen DQPM

In dit hoofdstuk worden de resultaten getoond van toepassing van de methode op de case die geselecteerd is voor dit onderzoek, stap 4 uit het plan van aanpak. Figuur 13 geeft een overzicht van de uitgevoerde activiteiten binnen deze stap en de samenhang daartussen.



Figuur 13 Stappen en samenhang stap 4, demonstratie

In de paragrafen hieronder worden de stappen uit de figuur toegelicht.

### 5.1. Voorbereiding

#### 5.1.1. Selecteren case

De geselecteerde case is het werkproces 'Controleren ingewonnen gegevens', onderdeel van het bedrijfsproces 'Beschikbaar stellen gegevens'. Dit proces beschrijft hoe het inwinnen van gegevens van derden zoals banken, verzekeringsmaatschappijen, multinationals en overheden van andere landen verloopt. De gegevens worden ontvangen, gecontroleerd, eventueel aangevuld en daarna beschikbaar gesteld voor interne en externe afnemers. Het proces voldoet aan de gestelde criteria: het bevat meerdere gegevensbewerkingen, door onder meer internationale wetgeving over gegevensuitwisseling tussen landen is het volop in ontwikkeling, er spelen diverse interne en externe stakeholders een rol en er hebben zich de afgelopen tijd verschillende problemen met datakwaliteit voorgedaan.

#### 5.1.2. Selecteren klankbordgroep

Om de leden van de klankbordgroep te kunnen vaststellen, is een stakeholderanalyse uitgevoerd. In eerste instantie is gekeken naar wie belang heeft bij procesmodellen, binnen die populatie wie belang heeft bij datakwaliteit. Stakeholders werden onderverdeeld naar belanghebbenden tijdens het informatievoorzienings (IV) traject, ná het IV traject en diverse niet IV gerelateerde stakeholders. Hierbij is gebruik gemaakt van eerdere inventarisaties door een interne kennisgroep. De stakeholders

zijn in een belangentabel gezet waarbij macht afgezet werd tegen belang. Dit leverde uiteindelijk de volgende samenstelling voor de klankbordgroep op (de namen zijn achterwege gelaten).

Tabel 7 Samenstelling klankbordgroep

| Kwadrant                           | Stakeholder                                     | Rol   |
|------------------------------------|---|---|
| <b>Belang hoog/<br/>Macht hoog</b> | Ontwikkelteams                                  | System engineer Agile Release train inwinnen  |
|                                    | CAP en kantoren                                 | Manager IVG CAP   |
|                                    | Business eigenaar                               | Product Owner Agile Release train inwinnen  |
|                                    | Leden project Dynamische Kwaliteitscontrole CAP | Kwaliteitsmedewerker  |
|                                    | Architecten agile release trains                | Architect Inning  |
|                                    | OAG   | Gegevensanalist   |
| <b>Belang laag/<br/>Macht hoog</b> | Business analisten                              | Business analyst Portfolioteam Gegevens   |
|                                    | Gebruikersgroep Architectuur en Ontwerp         | Voorzitter Gebruikersgroep Architectuur en Ontwerp                                  |
|                                    | COP O2  | Procesontwerper en Lid COP O2   |
| <b>Belang hoog/<br/>Macht laag</b> | Testers   | Testregisseur   |
|                                    | Schrijvers van werkinstructies en AO procedures | AO Medewerker   |
|                                    | Beheerders van ontwerpproducten IBS en STS      | Procesontwerper, ook lid COP O2 en schrijver Belastingdienst Conventiedocument BPMN |

Opvallend was dat door het methodisch aan te pakken de klankbordgroep een bredere en uitgebreidere samenstelling kreeg ten opzichte van een eerste brainstorm.

### 5.1.3. Vaststellen criteria

De criteria die in de literatuur zijn gevonden, zijn eerst geanalyseerd. Een aantal van de criteria bleek al geborgd te zijn in de interne standaarden en voorschriften van de onderzoeksorganisatie. In bijlage 5 is terug te vinden welke dat zijn en waar in de voorschriften deze terug te vinden zijn. De overige criteria zijn voorgelegd aan de klankbordgroep met het verzoek deze te wegen. Hiervoor is gebruik gemaakt van Analytic Hierarchy Process (AHP), een methode die het nemen van beslissingen en het wegen van criteria ondersteunt (Saaty, 1980). Bij AHP worden de criteria paarsgewijs tegen elkaar afgezet en gewogen waarmee een totaalscore gemaakt kan worden. Dit kan samengesteld worden uit de weging en score van meerdere deelnemers. Dit gaf het volgende beeld.

Tabel 8 weging criteria door klankbordgroep

| Decision Hierarchy           |   |           |
|------------------------------|---|-----------|
| Level 0                      | Level 1                                       | Glb Prio. |
| Datakwaliteit in procesmodel | Vertrouwd begrijpelijk niet complex 0.140     | 14.0%     |
|                              | Inzicht in bottlenecks 0.105                  | 10.5%     |
|                              | Bruikbaar als input systeemontwikkeling 0.064 | 6.4%      |
|                              | Zichtbaarheid dataflow 0.077                  | 7.7%      |
|                              | Transformatieproces duidelijk 0.106           | 10.6%     |
|                              | Zichtbaar waar DQ beïnvloed 0.096             | 9.6%      |
|                              | Zichtbaarheid kwaliteitscontroles 0.124       | 12.4%     |
|                              | Ondersteuning maatregelen DQ 0.066            | 6.6%      |
|                              | Meetbaarheid DQ product adhv dimensies 0.054  | 5.4%      |
|                              | Zichtbaarheid business impact 0.134           | 13.4%     |
|                              | Gedifferentieerde representatie data 0.034    | 3.4%      |
|                              |   | 1.0       |

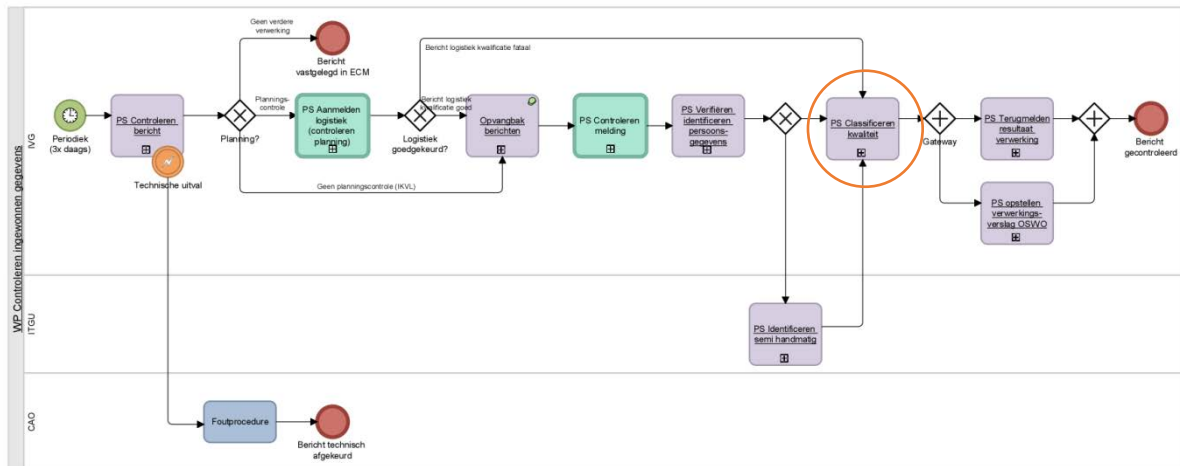
De reden om de in de literatuur gevonden criteria te wegen was om eventuele ontwerpkeuzes en mogelijke prioriteringskwesies in het traject te kunnen onderbouwen. Verder werd op deze manier de klankbordgroep al inhoudelijk betrokken.

## 5.2. Toepassing op de case

In dit hoofdstuk wordt verslag gedaan van het toepassen van de methode op de case, het bedrijfsproces 'Inwinnen verstrekken gegevens' met daarbinnen het werkproces 'Controleren ingewonnen gegevens'. Hierbij is de versie van de methode gebruikt van na de eerste iteratie.

### Stap 1, Beschrijf het informatieproduct

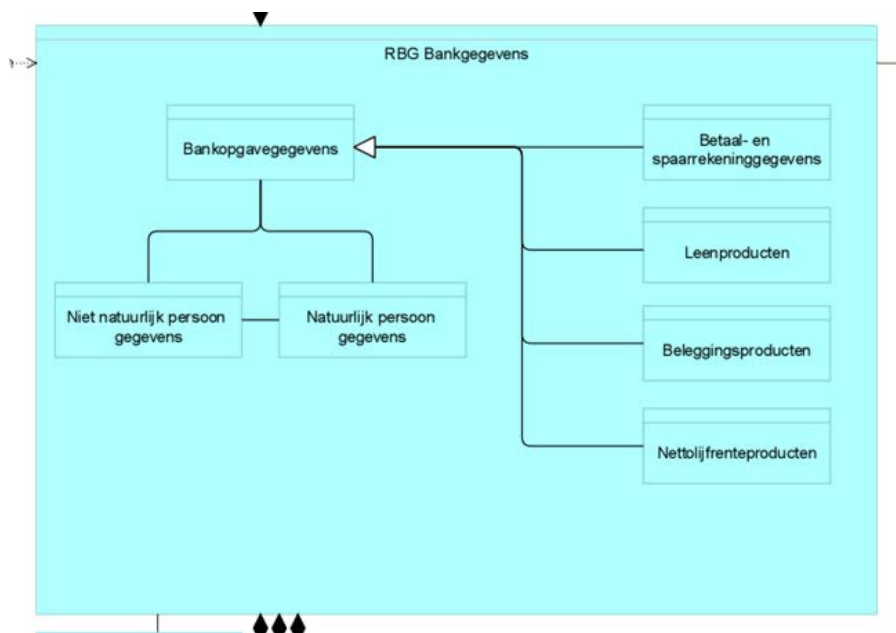
Het werkproces van de case is 'Controleren ingewonnen gegevens' en ziet er als volgt uit.



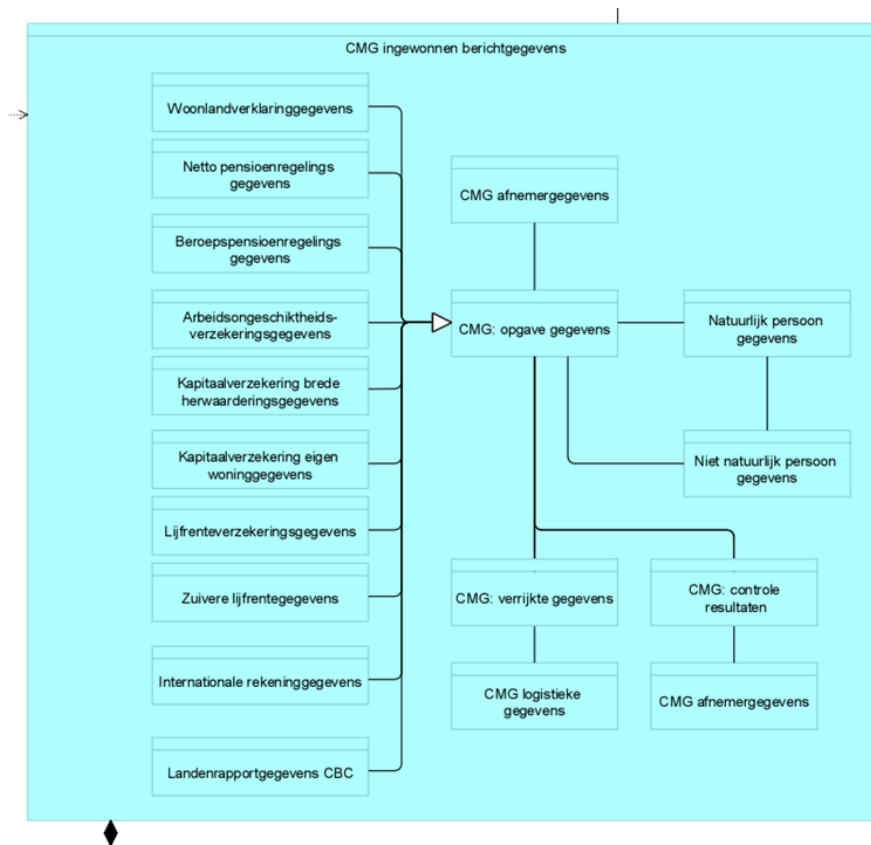
Figuur 14 Werkproces Controleren ingewonnen gegevens

Het informatieproduct dat door dit proces wordt geleverd is een gecontroleerd ingewonnen bericht dat beschikbaar gesteld kan worden aan interne en externe afnemers. Voorbeelden van dit product zijn een ingewonnen bankbericht waarin saldi van rekeninghouders zijn opgenomen voor diverse soorten rekeningen en een ingewonnen country by country bericht waarin gegevens van multinationals opgenomen zijn.

Voor dit product is niet één conceptueel gegevensmodel beschikbaar omdat er een variëteit aan berichten wordt ingewonnen die allemaal een andere inhoud hebben. Er is wel een data objectmodel beschikbaar waarin de ingewonnen gegevens als data-objecten zijn weergegeven. Uit het datamodel is alleen het deel opgenomen wat relevant is voor het informatieproduct van dit proces.



Figuur 15 Deel data-objectmodel RBG



Figuur 16 Deel van data-objectmodel CMG

De berichtgegevens worden één op één door geleverd, zonder bewerkingen of aanvullingen, met uitzondering van het burgerservicenummer (BSN) waar de ingewonnen gegevens betrekking op hebben. Dat nummer wordt gecontroleerd en in bepaalde berichtstromen opgezocht als het niet aanwezig is of niet correct is.

De afnemers zijn zowel intern als extern Belastingdienst. De belangrijkste interne afnemers zijn de VIA (de Vooraf Ingevulde Aangifte), de Inkomensheffingketen (IH) die de informatie als contra-informatie gebruikt bij het beoordelen van de aangifte en de afdeling Toezicht die gegevens raadpleegt als aangiften handmatig behandeld worden. Externe afnemers zijn het CBS en belastingdiensten van andere landen. De eisen die deze afnemers stellen zijn dat het bericht goedgekeurd is en geen fouten bevat.

In de omcirkelde processtap in figuur 14, PS Classificeren kwaliteit, is het informatieproduct gereed. Het product is dan gecontroleerd, eventueel aangevuld met BSN en de kwaliteit is vastgelegd. Daarna wordt alleen nog de afzender en een interne partij bij de Belastingdienst geïnformeerd over de kwaliteit van de aangeleverde gegevens.

#### Stap 2, Toets model op architectuurprincipes

In DQPM zijn de volgende principes verzameld. Per principe wordt aangegeven of het proces voldoet en zo nee, welke aanvullende requirements nodig zijn om aan het principe te voldoen. Hierbij moet het gehele bedrijfsproces in ogenschouw genomen worden, omdat veel zaken op het niveau van bedrijfsproces zijn ingericht, niet op het niveau van werkproces.

Tabel 9 Toetsing architectuurprincipes

| Principe   | Voldoet proces?   | Aanvullend requirement  |
|--|---|---|
| Quality at source  | Ja, de eerste twee werkprocessen in het bedrijfsproces zijn gericht op het controleren van het ontvangen bericht, te weten 'Ontvangen gegevens' waar gecontroleerd wordt of het bericht aan het geldende XSD formaat voldoet en 'Controleren ingewonnen gegevens' waarin op berichtniveau en op meldingniveau verschillende controles uitgevoerd worden.  | Niet van toepassing.  |
| Modelleer de gehele keten  | Ja, met twee kanttekeningen. Bepaalde processen zijn als blackboxpool weergegeven, deze zijn of in een ander domein verder uitgewerkt zoals bijvoorbeeld het werkproces 'Ontvangen gegevens' dat door Ontvangen/Mededelen verder beschreven en uitgewerkt is. Dat is voor het toetsen van dit principe geen bezwaar.<br>Een proces dat ontbreekt in het model is de analyse door, Onderzoek en Advies Gegevenskwaliteit, OAG. In de processtap 'Controleren bericht' worden gegevens verstuurd naar deze afdeling. Daar worden verschillende controles uitgevoerd, waarvan een aantal alleen maar op die plek, de resultaten daarvan teruggekoppeld aan de leveranciers van de gegevens.          | PR.1 Het procesmodel moet uitgebreid worden met het werkproces 'Analyseren kwaliteit gegevens' dat uitgevoerd wordt door OAG. Rationale: daarmee wordt het gehele ketenproces weergegeven.                                    |
| Breng de gebruiker naar de data, niet de data naar de gebruiker  | Op verschillende plekken in het proces wordt data gekopieerd en/of verplaatst: <ul style="list-style-type: none"> <li>Kopiëren van gegevens tussen werkproces 'Ontvangen gegevens' en 'Controleren ingewonnen gegevens'.</li> <li>Een kopie van het bericht naar OAG.</li> <li>Kopiëren van gegevens tussen werkproces 'Opslaan gegevens' en 'Verstrekken gegevens'.</li> </ul> Sommige van deze kopieslagen komen voort uit technische noodzaak, bijvoorbeeld bij OAG, waar de gegevens van een mainframe DB2 omgeving naar een SAS omgeving gekopieerd worden.<br>Van de andere kopiebewegingen is de impact te groot om in het kader van dit onderzoek aanvullende requirements op te stellen. | Niet van toepassing.  |
| Het besturingsproces is ingericht  | Ja, besturing op operationeel, tactisch en strategisch niveau is gemodelleerd.  | Niet van toepassing.  |
| Het informatieverwerkingsproces is volledig weergegeven (invoer, verwerking, bewaren, archiveren en vernietigen) | Archiveren en schonen is wel als werkproces onderkend, maar niet verder uitgewerkt.   | PR.2 Het werkproces Archiveren en schonen moet uitgewerkt worden voor de verschillende gevals categorieën die in het proces ingewonnen worden. Rationale: daarmee wordt het informatieverwerkingsproces volledig weergegeven. |
| Het informatieverwerkingsproces is zo efficiënt mogelijk ingericht (standaards en generieke diensten)            | Voor het ontvangen van berichten van leveranciers is gebruik gemaakt van standaard diensten van Ontvangen/Mededelen. Voor het verifiëren en identificeren van persoonsgegevens wordt gebruik gemaakt van diensten van   | Niet van toepassing.  |



| Principe  | Voldoet proces?   | Aanvullend requirement |
|---|---|------------------------|
|   | het systeem waarin de relaties worden beheerd, BVR.   |                        |
| Er zijn processen geïmplementeerd voor het corrigeren fouten in de throughput en output   | Ja, de controleresultaten van de gegevens worden vastgelegd. Fouten worden geclassificeerd met een F, fataal, verdere verwerking niet mogelijk, een C, verdere verwerking kan, maar fout moet gecorrigeerd worden, een S, fout hoeft niet gecorrigeerd te worden, maar aandachtspunt voor volgende levering. Middels een verwerkingsverslag wordt de leverancier op de hoogte gesteld. Voor fouten die halverwege ontstaan bij de logistieke controle is een foutprocedure ingericht. | Niet van toepassing.   |
| De raakvlakken van het onderzochte werkproces met andere processen zijn zichtbaar gemaakt in het model. Van de afnemende werkprocessen zijn de kwaliteitseisen beschreven die gesteld worden aan het informatieproduct. | In de procesarchitectuur zijn de afnemende processen zichtbaar gemaakt. In de bijbehorende beschrijving (interactieontwerp) is aangegeven welke eisen de afnemers stellen aan de gegevens. Aangegeven is welke foutkwalificatie de gegevens mogen hebben.   | Niet van toepassing.   |

### Stap 3, Bepaal relevante kwaliteitsdimensies

Voor deze case is het aantal kwaliteitsdimensies beperkt gehouden om het onderzoek niet te groot te laten worden. De volgende kwaliteitsdimensies worden in het onderzoek verder gebruikt:

Tabel 10 Overzicht gebruikte kwaliteitsdimensies case inwinnen

| Soort                 | Kwaliteitsdimensie       | Gebruik in case |
|-----------------------|--------------------------|-----------------|
| Intrinsieke kwaliteit | Juistheid                | Ja              |
|                       | Geloofwaardigheid        | Ja              |
|                       | Consistentie             | Nee             |
| Context               | Relevantie               | Nee             |
|                       | Tijdigheid               | Ja              |
|                       | Volledigheid             | Ja              |
| Representatie         | Interpreteerbaarheid     | Nee             |
|                       | Semantische consistentie | Nee             |
|                       | Begrijpelijkheid         | Nee             |
| Toegankelijkheid      | Toegankelijkheid         | Nee             |
|                       | Beveiliging              | Ja              |

De definities van deze kwaliteitsdimensies zoals deze in dit proces gebruikt zijn, zijn als volgt. Per dimensie is de gevals categorie opgenomen waar het in deze case om gaat.

Tabel 11 Definities kwaliteitsdimensies in case

| Kwaliteitsdimensie | Definitie   | Gevalscategorie                          |
|--------------------|---|--|
| Juistheid          | Batini et al. (2009) maken onderscheid tussen syntactische juistheid en semantische juistheid. Semantische juistheid geeft aan of een waarde <i>w</i> een afspiegeling vormt van de werkelijkheid. In deze case wordt alleen gekeken naar syntactische juistheid. Dat is de mate waarin een waarde <i>w</i> de elementen van een overeenkomstig definitiedomein <i>D</i> benadert.  | Bericht<br>Melding<br>Verwerkingsverslag |
| Geloofwaardigheid  | Moossavizadeh, Mohsenzadeh, & Arshadi (2012) definiëren geloofwaardigheid als de mate waarin data geaccepteerd als waar of als item dat waar, echt en geloofwaardig lijkt in een specifieke omgeving en in overeenstemming met relevante regels. Geloofwaardige data in een systeem kan of echt waar zijn en consistent met actuele data, of waar, echt en geloofwaardig lijken te zijn.  | Melding<br>Verwerkingsverslag            |
| Tijdigheid         | Batini et al. (2009) geeft in zijn inventarisatie van deze dimensie weer dat er drie aspecten zijn van tijdigheid: <ul style="list-style-type: none"><li>– Actualiteit, geeft aan hoe oud de informatie is ten opzichte van het moment dat het opgeslagen of gewijzigd werd. Van belang voor gegevens die snel verouderen.</li><li>– Tijdigheid, de mate waarin tijdig over gegevens beschikt kan worden.</li><li>– Volatiliteit, de (in)stabiliteit van informatie, de frequentie van wijzigingen in een waarde van een attribuut (Bovee et al. 2001) of de periode waarin gegevens overeenkomen met de waarde in de 'echte wereld' (Jarke et al. 1995).</li></ul> In deze case kijken we naar het tweede aspect, de mate waarin tijdig over gegevens beschikt kan worden. | Bericht<br>Melding<br>Verwerkingsverslag |
| Volledigheid       | Batini et al. (2009) geeft een overzicht van de verschillende definities in de literatuur van volledigheid. Ze variëren vooral in de context waarin volledigheid beoordeeld wordt; informatiesystemen, datawarehouses of entiteiten. Volledigheid   | Melding<br>Verwerkingsverslag            |

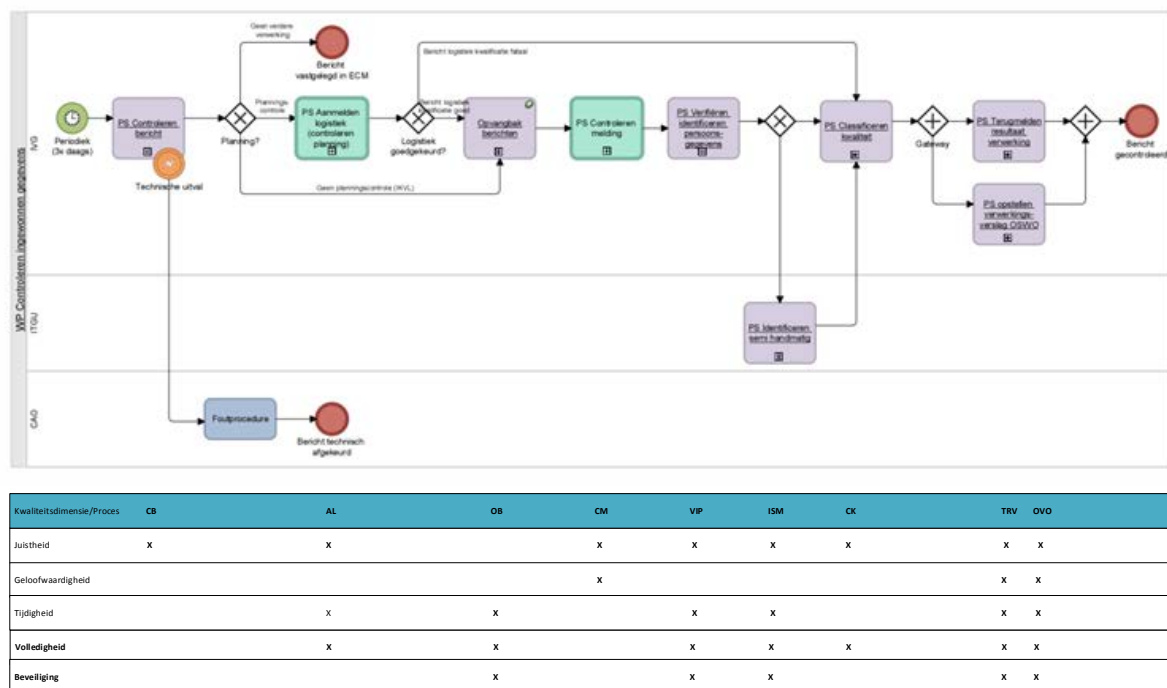
|             |   |                               |
|-------------|---|-------------------------------|
|             | wordt in deze case gezien als de mate waarin de dataelementen de totale relevante populatie vertegenwoordigen (dekingsgraad). |                               |
| Beveiliging | Beveiliging geeft aan in welke mate de gegevens beschermd zijn tegen ongeautoriseerde toegang (Y. W. Lee et al., 2002).       | Melding<br>Verwerkingsverslag |

Afgezet tegen de processtappen levert dat het volgende overzicht op.

Tabel 12 Tabelweergave processtappen afgezet tegen dimensies werkproces controleren ingewonnen gegevens

| Werkproces controleren ingewonnen gegevens | Controleren berichten | Aanmelden logistiek | Opvang bak berichten | Controleren Melding | IVP | Identif. semi handmatig | Class. kwaliteit | Terugm. resultaat verwerk. | Opstellen verw.versl. OSWO |
|--|-----------------------|---------------------|----------------------|---------------------|-----|-------------------------|------------------|----------------------------|----------------------------|
| Juistheid                                  | X                     | X                   |                      | X                   | X   | X                       | X                | X                          | X                          |
| Geloofw.                                   |                       |                     |                      | X                   |     |                         |                  | X                          | X                          |
| Tijdigheid                                 |                       | X                   | X                    |                     | X   | X                       |                  | X                          | X                          |
| Volledigheid                               |                       | X                   | X                    |                     | X   | X                       | X                | X                          | X                          |
| Beveiliging                                |                       |                     | X                    |                     | X   | X                       |                  | X                          | X                          |

In de grafische weergave ziet dat er als volgt uit.



Figuur 17 Grafische weergave processtappen afgezet tegen dimensies werkproces controleren ingewonnen gegevens

#### Stap 4, Beoordeel waarde kwaliteitsdimensies per stap – baseline situatie

De volgende bronnen zijn voor het uitvoeren van deze stap gebruikt:

- Procesmodel en procesarchitectuur inwinnen gegevens (laatste versie uit de repository).
- Ontwerpdocumentatie bij het proces (interactieontwerp, versie uit PI 9 2018).
- Functioneel ontwerpdokumentatie ontwikkelteams (versie PI 9 2018).

Voor deze stap wordt het beslismodel uitgevoerd die voor deze stap is opgesteld. Dit wordt gedaan per processtap uit het procesmodel. Onderstaande tabel laat de score zien, de rode regel geeft de score voor de eindbeslissing weer.

Tabel 13 Ingevuld DMN model beoordeling kwaliteitsdimensies per stap

|  | Informatie-item                                  | Controleren bericht      | Aanmelden logistiek      | Opvangbak berichten      | Controleren melding      | IVP                      |
|--|--|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Inputinformatie-items                            | Aanwezigheid documentatie                        | Aanwezig                 | Aanwezig                 | Aanwezig                 | Aanwezig                 | Aanwezig                 |
|  | Interviews stakeholders                          | Positief beeld           | Positief beeld           | Neutraal beeld           | Neutraal beeld           | Neutraal beeld           |
|  | Kwaliteitsmeting                                 | Positief beeld           | Positief beeld           | Neutraal beeld           | Neutraal beeld           | Neutraal beeld           |
|  | Uitwerking requirements DQ                       | Uitgewerkt of afleidbaar | Uitgewerkt of afleidbaar | Uitgewerkt of afleidbaar | Uitgewerkt of afleidbaar | Uitgewerkt of afleidbaar |
|  | Volledigheid kwaliteitsdimensies                 | Volledig                 | Volledig                 | Gedeeltelijk             | Gedeeltelijk             | Gedeeltelijk             |
| Afgeleide informatie-items:<br>afgeleide waarden | Datakwaliteit in procesuitvoering                | Goed                     | Goed                     | Redelijk                 | Redelijk                 | Redelijk                 |
|  | Score kwaliteitsdimensie voor BPMN concept       | Goed                     | Goed                     | Redelijk                 | Redelijk                 | Redelijk                 |
|  | Uitwerking requirements in model en documentatie | Goed                     | Goed                     | Redelijk                 | Redelijk                 | Redelijk                 |

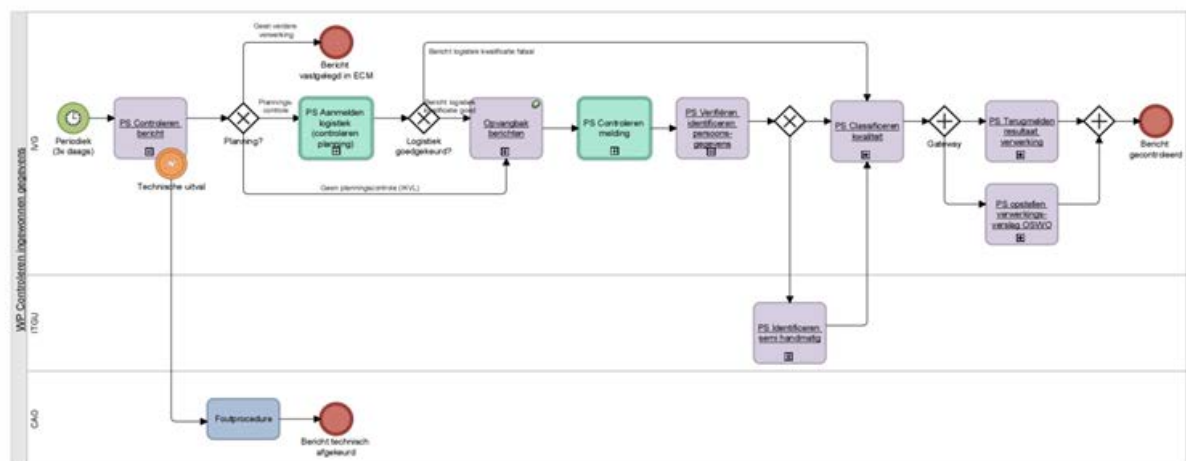
|  | Informatie-item                                  | Identificeren semi handm. | Classificeren kwaliteit            | Terugmelden resultaat verwerking   | Opstellen verwerkingsverslag OSWO  |
|--|--|---------------------------|------------------------------------|------------------------------------|------------------------------------|
| Inputinformatie-items                            | Aanwezigheid documentatie                        | Aanwezig                  | Aanwezig                           | Aanwezig                           | Aanwezig                           |
|  | Interviews stakeholders                          | Neutraal beeld            | Neutraal beeld                     | Neutraal beeld                     | Neutraal beeld                     |
|  | Kwaliteitsmeting                                 | Neutraal beeld            | Neutraal beeld                     | Neutraal beeld                     | Neutraal beeld                     |
|  | Uitwerking requirements DQ                       | Uitgewerkt of afleidbaar  | Niet uitgewerkt en niet afleidbaar | Niet uitgewerkt en niet afleidbaar | Niet uitgewerkt en niet afleidbaar |
|  | Volledigheid kwaliteitsdimensies                 | Gedeeltelijk              |                                    |                                    |                                    |
| Afgeleide informatie-items:<br>afgeleide waarden | Datakwaliteit in procesuitvoering                | Redelijk                  | Redelijk                           | Redelijk                           | Redelijk                           |
|  | Score kwaliteitsdimensie voor BPMN concept       | Redelijk                  | Slecht                             | Slecht                             | Slecht                             |
|  | Uitwerking requirements in model en documentatie | Redelijk                  | Slecht                             | Slecht                             | Slecht                             |

In de volgende tabel staat de score per kwaliteitsdimensie per processtap vastgesteld volgens de afleiding die de methode biedt.

Tabel 14 Tabelweergave score processtappen afgezet tegen dimensies werkproces controleren ingewonnen gegevens

| Werkproces controleren ingewonnen gegevens | Controleren berichten | Aanmelden logistiek | Opvang bak berichten | Controleren Melding | IVP | Identif. semi handmatig | Class. kwaliteit | Terugm. resultaat verwerk. | Opstellen verw.versl. OSWO |
|--|-----------------------|---------------------|----------------------|---------------------|-----|-------------------------|------------------|----------------------------|----------------------------|
| Juistheid                                  | +                     | +                   |                      | +                   | +   | +                       | +                | --                         | --                         |
| Geloofw.                                   |                       |                     |                      | --                  |     |                         |                  | --                         | --                         |
| Tijdigheid                                 |                       | +                   | --                   |                     | --  | --                      |                  | --                         | --                         |
| Volledigheid                               |                       | +                   | +                    |                     | --  | +                       | +                | --                         | --                         |
| Beveiliging                                |                       |                     | --                   |                     | --  | --                      |                  | --                         | --                         |

In de grafische weergave ziet dat er als volgt uit:



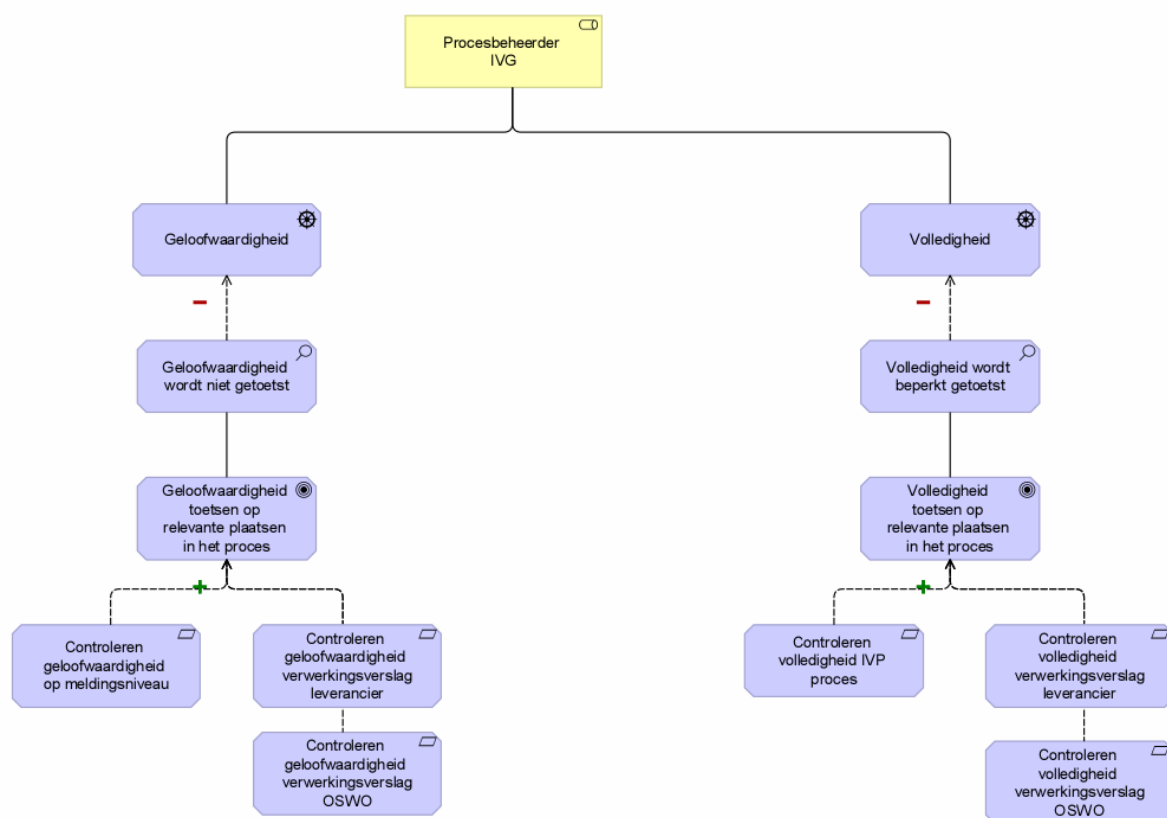
| Kwaliteitsdimensie/Proces | CB | AL | OB | CM | VIP | ISM | CK | TRV | OVO |
|---------------------------|----|----|----|----|-----|-----|----|-----|-----|
| Juistheid                 | 😊  | 😊  |    | 😊  | 😊   | 😊   | 😊  | 😞   | 😞   |
| Geloofwaardigheid         |    |    |    | 😞  |     |     |    | 😞   | 😞   |
| Tijdigheid                |    | 😊  | 😞  |    | 😞   | 😞   |    | 😞   | 😞   |
| Volledigheid              |    | 😊  | 😊  |    | 😞   | 😊   | 😊  | 😞   | 😞   |
| Beveiliging               |    |    | 😞  |    | 😞   | 😞   |    | 😞   | 😞   |

Figuur 18 Grafische weergave score processtappen afgezet tegen dimensies werkproces controleren ingewonnen gegevens

De tabel en de grafische weergave geven op twee manieren inzicht: welke dimensies goed of slecht scoren en welke processtappen. Concreet voor deze case betekent dit dat de dimensie juistheid een positief beeld geeft, op de meeste plekken in het proces scoort deze goed. Volledigheid is op veel plaatsen geborgd maar niet op alle. De andere dimensies verdienen aandacht. De eerste twee stappen in het proces scoren positief op de kwaliteitsdimensies, daarna scoort het meerendeel rood.

#### *Stap 5, Vaststellen requirements datakwaliteit*

De requirements zijn weergegeven in het volgende model met toelichting. Er is voor gekozen om de requirements als user story te formuleren, dat sluit beter aan bij de werkwijze van de onderzoeksorganisatie. De vorige stap liet zien dat er vanuit de kwaliteitsdimensies en vanuit de processtappen geredeneerd kan worden. Hier is de kwaliteitsdimensie gebruikt omdat deze als driver in het motivatiemodel opgenomen is. Alleen de dimensies geloofwaardigheid en volledigheid zijn uitgewerkt. Om de werking van het model te demonstreren hoeven niet alle dimensies of processtappen uitgewerkt te worden.



*Figuur 19 Uitwerking requirements voor twee dimensies*

*Tabel 15 Toelichting requirements voor twee dimensies*

| Requirement  | Toelichting  |
|--|--|
| <b>Controleren geloofwaardigheid op meldingsniveau</b> | <p>Als procesbeheerder van IVG wil ik meldingen kunnen controleren op geloofwaardigheid zodat ik de kwaliteit van de gegevens beter kan garanderen voor de afnemers.</p> <p>Geloofwaardigheid wordt bepaald door de volgende controles:</p> <ul style="list-style-type: none"> <li>het totaalbedrag van de meldingen voor de leverancier wijkt maximaal 20% af van wat er vorig jaar geleverd is.</li> <li>het aantal meldingen van de leverancier wijkt maximaal 20% af van wat vorig jaar</li> </ul> |

| Requirement   | Toelichting  |
|---|--|
|   | <p>geleverd is.</p> <ul style="list-style-type: none"> <li>het bedrag per melding wijkt maximaal 50% af van het gemiddelde voor dat berichttype voor alle leveranciers van vorig jaar.</li> </ul>  |
| <b>Geloofwaardigheid</b>  | De mate waarin data geaccepteerd als waar of als item dat waar, echt en geloofwaardig lijkt in een specifieke omgeving en in overeenstemming met relevante regels. Geloofwaardige data in een systeem kan of <u>echt</u> waar zijn en consistent met actuele data, of waar, echt en geloofwaardig <u>lijken</u> te zijn.   |
| <b>Controleren geloofwaardigheid verwerkingsverslag leverancier</b> | <p>Als procesbeheerder van IVG wil ik de geloofwaardigheid van de verstuurde verwerkingsverslagen voor de gegevensleverancier kunnen vaststellen zodat leveranciers goede verslagen krijgen.</p> <p>Geloofwaardigheid wordt bepaald door de volgende controles:</p> <ul style="list-style-type: none"> <li>het aantal fouten op het verslag is niet meer dan 150% van het aantal meldingen uit het bericht.</li> </ul>   |
| <b>Controleren geloofwaardigheid verwerkingsverslag OSWO</b>        | <p>Als procesbeheerder van IVG wil ik de geloofwaardigheid van de verstuurde verwerkingsverslagen voor OSWO kunnen vaststellen zodat OSWO goede verslagen krijgt.</p> <p>Geloofwaardigheid wordt bepaald door de volgende controles:</p> <ul style="list-style-type: none"> <li>het aantal fouten op het verslag is niet meer dan 150% van het aantal meldingen uit het bericht.</li> </ul>  |
| <b>Controleren volledigheid IVP proces</b>                          | <p>Als procesbeheerder van IVG wil ik de volledigheid van de IVP verwerking kunnen vaststellen zodat ik de kwaliteit beter kan garanderen voor afnemers.</p> <p>Volledigheid is in dit geval tweeledig:</p> <ul style="list-style-type: none"> <li>zijn alle meldingen die aan het IVP proces verstuurd worden ook weer terugontvangen?</li> <li>zijn alle meldingen voorzien van een controleresultaat?</li> </ul>  |
| <b>Volledigheid</b>   | <p>Volledigheid kan gezien worden als de mate waarin de dataelementen de totale relevante populatie vertegenwoordigen (dekingsgraad). Volledigheid kan ook gezien worden als de mate waarin alle velden in een regel, kolom of tabel gevuld zijn met een relevante waarde (vullingsgraad). In beide gevallen is belangrijk om te weten waarom een waarde niet gevuld is:</p> <ul style="list-style-type: none"> <li>Waarde bestaat niet (iemand heeft geen mobiel nummer)</li> <li>Waarde bestaat wel, maar is niet bekend (onbekend of iemand een mobiel nummer heeft)</li> <li>Onbekend of waarde bestaat</li> </ul> |
| <b>Controleren volledigheid verwerkingsverslag leverancier</b>      | <p>Als procesbeheerder van IVG wil ik de volledigheid van de verwerkingsverslagen voor de leverancier kunnen vaststellen zodat leveranciers goede verslagen krijgen.</p> <p>Volledigheid is in dit geval tweeledig:</p> <ul style="list-style-type: none"> <li>zijn alle meldingen die verwerkt zijn voor die specifieke leverancier opgenomen op het verwerkingsverslag?</li> <li>hebben alle leveranciers een verwerkingsverslag ontvangen?</li> </ul>   |
| <b>Controleren volledigheid verwerkingsverslag OSWO</b>             | <p>Als procesbeheerder van IVG wil ik de volledigheid van de verwerkingsverslagen voor OSWO kunnen vaststellen zodat OSWO goede verslagen krijgt.</p> <p>Volledigheid is in dit geval tweeledig:</p> <ul style="list-style-type: none"> <li>zijn alle meldingen die verwerkt zijn voor die specifieke leverancier opgenomen op het verwerkingsverslag?</li> <li>hebben alle leveranciers een verwerkingsverslag ontvangen?</li> </ul>  |

#### Stap 6, Toevoegen requirements aan procesmodel - targetsituatie

Onderstaande tabel geeft de relatie weer tussen de opgestelde requirements en de stappen uit het proces waar deze betrekking op hebben.

Tabel 16 Koppeling requirements aan processtappen

|  |  | IVGD stap 5 requirements | Motivatie-elementen | Controleren geloofwaardigheid op meldingsniv | Controleren geloofwaardigheid verwerkingsvers | Controleren geloofwaardigheid verwerkingsvers | Controleren volledigheid identificeren semi han | Controleren volledigheid IVP proces | Controleren volledigheid opvangbak | Controleren volledigheid verwerkingsverslag lev | Controleren volledigheid verwerkingsverslag OS | Geloofwaardigheid | Geloofwaardigheid toetsen op relevante plaatse | Geloofwaardigheid wordt niet getoetst | Volledigheid | Volledigheid toetsen op relevante plaatsen in h | Volledigheid wordt beperkt getoetst |
|--|--|--------------------------|---------------------|--|---|---|---|-------------------------------------|------------------------------------|---|--|-------------------|--|---------------------------------------|--------------|---|-------------------------------------|
|  | PS Aanmelden logistiek (controleren ...  |                          |                     |  |   |   |   |                                     |                                    |   |  |                   |  |                                       |              |   |                                     |
|  | PS Classificeren kwaliteit               |                          |                     |  |   |   |   |                                     |                                    |   |  |                   |  |                                       |              |   |                                     |
|  | PS Controleren bericht                   |                          |                     |  |   |   |   |                                     |                                    |   |  |                   |  |                                       |              |   |                                     |
|  | PS Controleren melding                   |                          |                     | X  |   |   |   |                                     |                                    |   |  |                   |  |                                       |              |   |                                     |
|  | PS opstellen verwerkings-verslag OS...   |                          |                     |  |   | X   |   |                                     |                                    |   | X  |                   |  |                                       |              |   |                                     |
|  | PS Terugmelden resultaat verwerking      |                          |                     |  | X   |   |   |                                     |                                    | X   |  |                   |  |                                       |              |   |                                     |
|  | PS Verifiëren identificeren persoons-... |                          |                     |  |   |   |   | X                                   |                                    |   |  |                   |  |                                       |              |   |                                     |
|  | Lane set                                 |                          |                     |  |   |   |   |                                     |                                    |   |  |                   |  |                                       |              |   |                                     |

### 5.3. Verzamelen feedback klankbordgroep

De methode is in één iteratie ontwikkeld, waarbij het verzamelen van feedback de iteratie onderverdeeld is in drie fases. De initieel ontwikkelde methode is in de eerste fase besproken met de afstudeerbegeleider. Vooruitlopend op de klankbordgroep is in de tweede fase een proefsessie gehouden met een drietal collega's die geselecteerd zijn op hun expertise. Doel van de proefsessie was om te beoordelen of de opzet van de bijeenkomst zou voldoen en of de vragenlijst aanpassingen behoefde. De initiële opzet voldeed waarna de klankbordgroep in de derde fase – in twee deelsessies – de methode beoordeeld heeft.

Zo'n sessie bestond uit een aantal delen. Van te voren kregen de deelnemers de samenvatting van het literatuuronderzoek toegestuurd. Tijdens de bijeenkomst werd het doel en het belang van het onderzoek toegelicht waarna de methode stap voor stap toegelicht werd. Daarna werd getoond hoe de methode werkt als hij toegepast wordt op een case, het proces inwinnen en controleren van gegevens. Gedurende de hele bijeenkomst was er gelegenheid tot het stellen van vragen. De verzamelde feedback in dit gedeelte was heel divers. Het doel was vooral om het eerste criterium te toetsen, de bruikbaarheid en de duidelijkheid van de methode, maar er kwamen ook punten in relatie tot de andere criteria naar boven.

De vragen in de enquête werden na afloop buiten de bijeenkomst en individueel ingevuld. De enquête was bedoeld om alle criteria te toetsen. Alle feedback van zowel de bijeenkomst als de vragenlijst is verzameld in een tabel. Per regel is de feedback geanalyseerd en is bepaald of aanpassing van de methode nodig was en waarom.

### 5.3.1. Feedback klankbordgroep

Alle feedback is verzameld in een tabel waarin de volgende kolommen opgenomen zijn:

- Bron (begeleider, proefbijeenkomst, klankbordgroep).
- Tijdstip (sessies zijn opgenomen, tijdstip verwijst naar moment in de opname).
- Volgnummer.
- Bevinding/vraag/opmerking, een samenvatting van de bevinding.
- Verwerking, een analyse van de bevinding en wat de impact is op de methode.
- Aanpassing methode, of 'Nee' of 'Ja' en een verwijzing naar de stap met in steekwoorden de benodigde aanpassing.

Hieronder een fragment uit de tabel. De volledige tabel is opgenomen in bijlage 7.

| Bron | Tijdstip  | Nr.   | Bevinding/vraag/opmerking   | Verwerking  | Aanpassing methode  |
|------|---|-------|---|---|---|
| FG1  | 1:02:30<br>1:19:02                                  | FG.12 | Wees bewust van wat rode <b>smiley's</b> doen met managers, die willen alles op groen hebben. Als je in het traject bepaalde risico's accepteert, laat die dan niet rood, maar geef die een andere kleur. Wit bijvoorbeeld.             | Meenemen in de aanpak, stap 7, fase van selecteren welke risico's je af wilt dekken en welke je accepteert. Geaccepteerde risico's wel zichtbaar maken maar met afwijkende kleur.   | Ja, stap 7 splitsen, en selectie toevoegen. Zie ook FG.5.       |
| FG1  | 1:06:22   | FG.13 | Het procesmodel van deze case is een generiek model. Er lopen meerdere gevals categorieën doorheen. In hoeverre is de gevalscategorie van invloed op de werking van het model?  | Bij bepalen van de scope kiezen waar model op toegepast gaat worden. Zie ook FG.11.   | Ja, inleiding uitbreiden over scope.                            |
| FG1  | 1:06:22<br>1:28:30<br>1:30:32<br>1:32:09<br>1:09:57 | FG.14 | Meer bronnen nodig dan alleen procesmodel. Is procesmodel weergave van de werkelijkheid? Is het voor interviews nodig om stakeholders vast te stellen?  | Klopt, ook interviews etc. Toevoegen aan beslismodel stap 4. Ook meenemen: cijfers over de productie, bijvoorbeeld aantal incidenten en aantal verstoringen. Metingen uit <b>Solunk</b> . Mogelijk stakeholders in kaart brengen. | Ja, stap 4, toevoegen meerdere bronnen. Stakeholders toevoegen. |
| FG1  | 1:08:16<br>1:15:55                                  | FG.15 | Volledigheid bij processtap opvangbak is een andere volledigheid dan bij de processtap logistiek. Hoe kan je dat laten zien. En kun je volledigheid geïsoleerd zien of moet je het afzetten ten opzichte van een cijfer van vorig jaar. | Zie ook FG.8, definities toevoegen bij stap 3 en dat ook per processtap doen.   | Ja, stap 3, definities toevoegen, per processtap.               |
| FG1  | 1:15:01   | FG.16 | In de uitwerking van de case klopt er iets inhoudelijk niet. In het requirement staat dat het voor een melding geldt, maar het geldt voor totalen van de meldingen.   | Klopt, uitwerking aanpassen.  | Nee, niet methode, maar wel toepassing in case.                 |
| FG1  | 1:23:03   | FG.17 | Stap 6 helpt om het effect van je maatregelen te bepalen.   | Meenemen in rationale van stap 6.   | Ja, stap 6, rationale uitbreiden.                               |
| FG1  | 1:23:47   | FG.18 | Op welke processen pas je deze methode toe? Wanneer is het  | Zie ook FG.11 en FG.13.   | Ja, inleiding uitbreiden  |

Figuur 20 Fragment tabel met feedback klankbordgroepen

De belangrijkste bevindingen per stap zijn hieronder samengevat.

|  |  |
|--|--|
| Inleiding en uitgangspunten                      | <ul style="list-style-type: none"> <li>• Aannames en uitgangspunten expliciet gemaakt, bijvoorbeeld veronderstelde voorkennis bij stakeholders en aanwezigheid van bepaalde producten.</li> <li>• Definities van procesdecompositie toegevoegd, zie stap 2.</li> </ul>   |
| Stap 1, beschrijf het informatieproduct          | <ul style="list-style-type: none"> <li>• In eerste instantie gaf deze stap als resultaat een markering in het proces waar data gebruikt wordt. Bij toepassing bleek dat deze stap geen toegevoegde waarde bood omdat dit bijna overal plaatsvindt. De stap is vervangen door het beschrijven van het informatieproduct, waar het geproduceerd wordt in het proces, de betekenis van de gegevens, een inventarisatie van de afnemers en de eisen die ze stellen.</li> </ul> |
| Stap 2, toets model op architectuurprincipes     | <ul style="list-style-type: none"> <li>• Er zijn een aantal principes toegevoegd.</li> <li>• Voor deze stap verschuift het aandachtsgebied van werkproces naar bedrijfsproces of ketenproces. Deze constatering is toegevoegd. In de uitgangspunten zijn de definities van de verschillende lagen in procesdecompositie toegevoegd.</li> </ul>   |
| Stap 3, bepaal relevante kwaliteitsdimensies     | <ul style="list-style-type: none"> <li>• Diverse cosmetische zaken die de leesbaarheid van de producten uit de methode ten goede komen.</li> <li>• Definities toevoegen in de context van het proces waarin de methode toegepast wordt.</li> </ul>   |
| Stap 4, beoordeel waarde kwaliteitsdimensies per | <ul style="list-style-type: none"> <li>• In de beoordeling niet alleen het procesmodel betrekken, maar ook andere bronnen, bijvoorbeeld metingen in het proces, interviews met stakeholders en aantal gemelde incidenten.</li> </ul>   |



|   |  |
|---|--|
| stap – baseline situatie  | <ul style="list-style-type: none"> <li>• Het beslismodel is uitgebreid om deze extra bronnen mee te nemen in de beoordeling van de kwaliteitsdimensie.</li> <li>• Het beslismodel is gemodelleerd in Decision Model and Notation, DMN, een wereldwijde standaard om beslissingen te modelleren en beheerd door de Object Management Group, de organisatie die ook BPMN beheert, de procesmodelleertaal die in deze methode gebruikt wordt.</li> <li>• Het organiseren van het scoren is uitgewerkt.</li> </ul> |
| Stap 5, vaststellen requirements datakwaliteit                  | <ul style="list-style-type: none"> <li>• Tekstuele verduidelijking van de stap.</li> </ul>   |
| Stap 6, toevoegen requirements aan procesmodel – targetsituatie | <ul style="list-style-type: none"> <li>• Tekstuele verduidelijking van de stap.</li> </ul>   |
| Stap 7, Realiseren requirements                                 | <ul style="list-style-type: none"> <li>• Stap verder uitgewerkt. Voordat requirements gerealiseerd gaan worden zal er eerst geprioriteerd worden en een keuzeprocess plaatsvinden. Bepaalde risico's worden mogelijk geaccepteerd. Dit wordt zichtbaar gemaakt in het model.</li> </ul>  |

## 6. Evaluatie DQPM

Het doel van deze stap is om te beoordelen of de nieuwe methode de organisatie daadwerkelijk handvatten geeft om te beoordelen of de beïnvloeding van datakwaliteit in het informatieverwerkende proces zichtbaar gemaakt kan worden om daarmee vanuit een procesperspectief de kwaliteit van data te verbeteren. Tijdens de bijeenkomsten met de klankbordgroep zijn opnames gemaakt met een toepassing waarbij ook aantekeningen gemaakt kunnen worden gemaakt op specifieke momenten in de tijd. Dat is gebruikt om de feedback gericht te kunnen verzamelen en verwerken. De feedback is daarom niet getranscribeerd zoals in eerste instantie opgenomen was in het plan van aanpak. Condensatie heeft plaats gevonden in de tabel zoals in de vorige paragraaf besproken.

Daarnaast is een vragenlijst afgenomen bij de klankbordgroep om het model op toepasbaarheid en effectiviteit te beoordelen. De vragenlijst en de uitkomst per vraag zijn opgenomen in bijlage 8 en bijlage 9. De vragenlijst is zodanig opgesteld dat de methode op alle criteria waar de methode aan moet voldoen beoordeeld wordt. In deze paragraaf worden de belangrijkste bevindingen weergegeven. Onderstaande tabel laat het criterium zien en het resultaat naar aanleiding van de vragenlijst met een conclusie ten aanzien van het voldoen aan het criterium. Over het algemeen scoort de methode goed. Op een aantal punten is een volgende iteratie of verder onderzoek nodig.

Tabel 17 Criteria afgezet tegen resultaten vragenlijst

| Criterium  | Resultaat n.a.v. vragenlijst  |
|--|---|
| Modellen moeten vertrouwd zijn voor stakeholders, begrijpelijk zijn en niet te complex.<br>Weging 14%              | In de vragenlijst is uitgevraagd of de stappen duidelijk zijn en of duidelijk is wat de methode op gaat leveren. Voor de belangrijkste stappen is de duidelijkheid en de toegevoegde waarde nog eens extra getoetst. Bij iedere vraag was de meerderheid het eens met de vraag of stelling. De conclusie is dat de methode voldoet aan dit criterium.   |
| Het model moet inzicht geven in bottlenecks in het proces.<br>Weging 10,5%   | Een procesmodel laat de stappen zien die nodig zijn om een informatieproduct te maken. Bij de onderzoeksorganisatie is het verplicht om aantallen en doorlooptijd per stap op te geven. Dat helpt om een beeld te krijgen van bottlenecks. Echter, voor een beter beeld is het ook nodig om naar de uitvoering van het proces te kijken. Roser, Lorentzen & Deuse (2014) bijvoorbeeld nemen het observeren van processen en voorraadpunten mee in hun analyse van bottlenecks.<br>Conclusie: Voor deze methode is het zichtbaar maken van bottlenecks voldoende geborgd. Ligt het accent van een procesanalyse op het vinden van bottlenecks, dan zijn aanvullende stappen nodig zoals het observeren van procesuitvoering en voorraadpunten. |
| De opgeleverde procesmodellen moeten bruikbaar zijn als input voor het systeemontwikkelingstraject.<br>Weging 6,4% | Dit criterium is in een aantal vragen getoetst. Over het algemeen scoort dit criterium voldoende, maar er zijn een aantal kleine aanpassingen in stap 7 nodig. En verder levert dit punt een item op voor vervolgonderzoek, de mate waarop de requirements uit stap 5 aansluiten bij de agile voortbrenging van de onderzoeksorganisatie.   |
| Het model moet de raw dataflow, component dataflow en IP-flow kunnen laten zien.<br>Weging 7,7%                    | Dit criterium scoort voldoende, geen aanpassing nodig van de methode voor dit punt.   |
| Het transformatieproces moet duidelijk zijn.<br>Weging 10,6%   | De meerderheid is niet eens met deze stelling.<br>Vermoedelijk komt dat doordat het procesmodel in lagen op gebouwd is. Het model dat in het onderzoek gebruikt is geeft het proces weer op processtapniveau. Op dat niveau is de gegevensbewerking niet zichtbaar. Dat is pas zichtbaar op het niveau van activiteit. Dit niveau is met de deelnemers van de klankbordgroep niet besproken.<br>Conclusie: deze bevinding heeft verder onderzoek. Als er tijd is in een volgende iteratie meenemen, anders opnemen als punt voor  |

|   |  |
|---|--|
|   | vervolgonderzoek. Het gewicht van het criterium geeft aan dat de klankbordgroep dit een belangrijk punt vindt.   |
| De belangrijkste fasen in het proces en kritieke fasen die datakwaliteit beïnvloeden moeten gevisualiseerd kunnen worden.<br>Weging 9,6%  | Dit criterium scoort voldoende, geen aanpassing nodig van de methode voor dit punt.  |
| Kwaliteitscontroles zijn expliciet zichtbaar.<br>Weging 12,4%   | Dit criterium scoort voldoende, geen aanpassing nodig van de methode voor dit punt.  |
| Het model moet ondersteunend zijn in het implementeren van kwaliteitsmaatregelen.<br>Weging 6,6%  | Dit criterium scoort voldoende, geen aanpassing nodig van de methode voor dit punt   |
| Het model moet het mogelijk maken om op verschillende abstractieniveaus de kwaliteit van het informatieproduct meetbaar te maken aan de hand van de relevante kwaliteitsdimensies.<br>Weging 5,4% | Een groot deel geeft neutraal aan bij deze vraag, een minderheid is het eens is met deze stelling maar niemand is het niet eens met de stelling. Analyse: In de open reacties wordt verder geen toelichting gegeven. Het criterium scoort niet overtuigend. Tijdens de klankbordbijeenkomsten kwam naar voren dat het proces uit het onderzoek niet helemaal representatief was in de zin dat er geen 'echt' product gemaakt werd, maar meer een halffabrikaat dat door andere processen afgenomen wordt en verder bewerkt. Conclusie: ondanks het lage gewicht van het criterium en de neutrale tot positieve score, toch opnemen als punt voor verder onderzoek. Het raakt wel de essentie van de methode. Zie ook de analyse bij het volgende criterium.  |
| De methode moet concepten bevatten om het belang van de business impact en waarde van datakwaliteit te kunnen laten zien.<br>Weging 13,4%   | Voor dit punt is minder dan 50% het eens of helemaal eens met de vraag. Een derde is het oneens en neutraal. Analyse: gezien de score en het gewicht van het criterium behoeft dit punt aandacht. Vermoedelijk komt deze score doordat de scope van het onderzoek zich beperkt heeft tot het niveau van werkproces. Producten en diensten worden voortgebracht op het niveau van bedrijfsproces of ketenproces en door naar dat niveau te kijken kan de impact op de business beter beoordeeld worden. Uit de klankbordbijeenkomsten is al meegenomen dat de relatie van het werkproces in de context van bedrijfs- en ketenproces moet worden meegenomen. Conclusie: punt is meegenomen in de bevindingen bij stap 2 van de methode. Toetsen in de volgende iteratie of vervolgonderzoek of het met deze aanpassing beter scoort. |
| De methode moet rekening houden met de relevante context van data zodat een gedifferentieerde representatie van data mogelijk is.<br>Weging 3,4%  | Een meerderheid antwoordt positief of neutraal op deze vraag. Een kwart is het oneens met de stelling. Analyse: zie ook de vermoedelijke oorzaak zoals genoemd bij het criterium transformatieproces Het procesmodel is in lagen op gebouwd. Het model dat in het onderzoek gebruikt is geeft het proces weer op processtapniveau. Op dat niveau is de gegevensbewerking niet zichtbaar. Dat is pas zichtbaar op het niveau van activiteit. Dit niveau is met de deelnemers van de klankbordgroep niet besproken. Conclusie: het criterium scoort laag op belangrijkheid. De vermoedelijke verklaring vraagt niet direct om de methode aan te passen. Als in de volgende iteratie activiteitsniveau wordt meegenomen dan de vraag opnieuw evalueren of anders meenemen in vervolgonderzoek.  |

## 7. Conclusie, discussie en aanbevelingen, reflectie

In dit hoofdstuk worden de conclusies uit het onderzoek gegeven. Er volgen aanbevelingen voor de praktijk en voor vervolgonderzoek. Ook wordt ingegaan op methodologische issues zoals die in het plan van aanpak zijn onderkend.

### 7.1. Discussie

De opdracht in dit onderzoek was te onderzoeken hoe een datakwaliteitsgeoriënteerde manier van procesmodelleren stakeholders in een organisatie kan helpen om grip te krijgen op datakwaliteit om daarmee de kwaliteit van data te verhogen. Kijkend naar de resultaten van de klankbordgroep en de enquête lijkt het er in eerste instantie op dat dit doel gehaald is. Het model is duidelijk en begrijpelijk voor de stakeholders, het laat zien waar bottlenecks zitten, het laat zien waar problemen met datakwaliteit zich voordoen en het helpt om maatregelen te ontwikkelen om problemen op te lossen. In de paragraaf met aanbevelingen voor verder onderzoek staan nog wel een aantal aanbevelingen om de methode verder te ontwikkelen. De methode bevat de fasering die als rode draad uit het onderzoek van Batini et al. (2009) naar voren kwam; een beschrijvende fase, een assessment en een verbeterfase.

Het literatuuronderzoek begon met een paragraaf hoe datakwaliteit bepaald wordt. In de literatuur zijn verschillende uitgangspunten te vinden waarvoor dit onderzoek gekozen is voor die van Yang et al. (2013) (datakwaliteit is de mate waarin data bruikbaar is voor het gegeven doel in een gegeven context) omdat die het meest omvattend leek. In de loop van het onderzoek en tijdens de bijeenkomsten van de klankbordgroepen bleek dit de juiste keuze te zijn. Uit de sessie kwam terug dat de context van belang is en dat zelfs de definitie van een kwaliteitsdimensie per processtap per gevalscategorie een andere betekenis kan hebben.

Een interessante bevinding is dat in de literatuur het niet gebruiken van gegevens als kernprobleem wordt gezien van slechte datakwaliteit (Strong et al., 1997) terwijl in het onderzoek bleek dat de gevolgen van het gebruik van gegevens van slechte kwaliteit als veel ernstiger werd ervaren. Voor de onderzoeksorganisatie geldt dat gevolg- en imagoschade door de grote aantallen enorm kunnen zijn. Een fout in 1% van de ontvangen gegevens van banken die in de Vooraf Ingevulde Aangifte terecht komen, treft tienduizenden burgers. Alle inspanningen van de onderzoeksorganisatie ten aanzien van datakwaliteit zijn gericht op het voorkomen van problemen met datakwaliteit. Opvallend is wel dat de datageoriënteerde manier daarbij vaker gehanteerd wordt dan de procesgeoriënteerde manier. De methode DQPM die in dit onderzoek ontwikkeld is gaat uit van de procesbenadering. Batini et al. (2009) geven binnen die benadering nog twee richtingen: beheersing van het proces, een meer reactieve benadering, en herontwerp van het proces waarmee de oorzaken weggenomen worden. Afhankelijk van de uitkomst van stap 5, waarin de requirements vastgesteld worden om te verbeteren, kan DQPM beide richtingen aan, beheersing of herontwerp of een combinatie van beide.

In het literatuuronderzoek is een beeld gegeven van de ontwikkeling van datakwaliteit in procesmodellen. Methoden die daar aan bijdragen moeten aan drie belangrijke eisen voldoen: ze moeten aansluiten bij standaard Business Process Management (BPM) methoden en begrijpelijk zijn voor stakeholders, ze moeten op een systematische manier laten zien hoe een informatieproduct voortgebracht wordt en welke organisaties en systemen daarbij betrokken zijn en het model moet ondersteunend zijn om kwaliteitsmaatregelen te implementeren. Veel methoden uit de literatuur

voldoen maar aan één van die criteria. IP Map bijvoorbeeld van Shankaranarayanan et al. (2000) laat bijvoorbeeld zien hoe een informatieproduct wordt voortgebracht en welke controles in het proces zitten, maar is geen standaard BPM methode en is niet begrijpelijk voor stakeholders. Ofner et al. (2012) gebruikt wel een standaard BPM methode, maar maakt het met business rules erg complex voor stakeholders en gaat er daarnaast van uit dat kwaliteit te meten is. Voor nieuwe processen geldt dit nog niet terwijl ook daar behoefte is aan kwaliteitszorg. Bij de uitvoering van het onderzoek kwam de aanpak van Rodríguez et al. (2012) in beeld, die sluiten ook aan bij een standaard BPM methode, maar hebben het metamodel vervolgens zodanig aangepast dat het voor veel organisaties niet toepasbaar zal zijn. DQPM heeft alleen een visualisatie toegevoegd die met een kantoorautomatiseringshulpmiddel gerealiseerd kan worden. Ook is aansluiting gezocht met andere standaarden als DMN en Archimate om de acceptatie bij stakeholders te vergroten. Bij het ontwikkelen van DQPM als methode is geprobeerd om aan alle criteria te voldoen. Kijkend naar de resultaten van het onderzoek dan is dit op het eerste gezicht gelukt. De methode voldoet aan de criteria. Op een aantal punten is echter nog verder onderzoek nodig om te toetsen of voorgestelde aanpassingen het verwachte effect hebben. Pas dan kan met meer zekerheid gezegd worden of DQPM een substantiële stap is in een procesbenadering van datakwaliteit.

## 7.2. Conclusies

Het doel van het uitgevoerde onderzoek was om organisaties handvatten te geven problemen met datakwaliteit, die vaak technisch van aard zijn, zichtbaar te maken, in business termen begrijpelijk te maken en te kunnen analyseren en op te kunnen lossen in zowel de ontwerpfase als de uitvoeringsfase van een proces.

Om dit doel te bereiken is onderzocht of een datakwaliteitgeoriënteerde manier van procesmodelleren bruikbaar is voor stakeholders in een organisatie om problemen met datakwaliteit op te sporen en inzichtelijk te maken. Daartoe is eerst onderzocht hoe de kwaliteit van data bepaald wordt, hoe kwaliteit inzichtelijk gemaakt kan worden door middel van dimensies en welke factoren datakwaliteit beïnvloeden. Er is onderzocht hoe die factoren zichtbaar gemaakt kunnen worden in een procesmodel. Bestaande procesmodelleertalen bieden daar geen hulpmiddelen voor. In de literatuur zijn alternatieven te vinden die echter niet aansluiten bij standaard BPM methoden. In dit onderzoek is aansluiting gezocht bij standaard BPM methoden. Er is een beslismodel opgesteld om de waarde van een dimensie te kunnen specificeren, ook als er geen meetgegevens beschikbaar zijn. Verder is een visualisatie ontwikkeld om de waarde van kwaliteitsdimensies voor een specifiek proces in een specifieke context te plotten op een proces gemodelleerd in BPMN, een standaard modelleertaal voor processen. In het onderzoek is getoetst of de visualisatie de stakeholders helpt om problemen met datakwaliteit inzichtelijk te maken en op te lossen. De uitkomst daarvan was positief, de modellen zijn inzichtelijk en vertrouwd voor de stakeholders, bottlenecks worden zichtbaar, de factoren die de kwaliteit van data beïnvloeden zijn zichtbaar en de uitkomsten zijn bruikbaar voor het IV voortbrengingstraject.

Ook de andere elementen van de onderzoeksdoelstelling zijn met behulp van de literatuur vertaald naar meetbare criteria die getoetst zijn in de onderzoeksorganisatie. De methode is zowel bruikbaar voor bestaande processen waar al meer te vinden is (en te meten) over datakwaliteit als voor nieuwe processen. Wel kwam uit het onderzoek naar voren dat niet alle problemen opgelost hoeven te worden. Door allerlei factoren in de organisatie (beschikbare capaciteit en budget) kunnen

bepaalde risico's geaccepteerd worden. Ook dit kan zichtbaar gemaakt worden in DQPM door een afwijkende kleur van de score van de dimensie in het proces.

De basis om de waarde van een kwaliteitsdimensie te scoren is in de loop van het onderzoek verbreed. In eerste instantie werd alleen het procesmodel en eventuele ontwerpdocumentatie gebruikt. Daar zijn later ook andere beschikbare bronnen aan toegevoegd zoals interviews, meetgegevens en incidentrapportages. Oplossingen worden niet alleen zichtbaar gemaakt in het procesmodel, maar ook gemodelleerd in de motivatie-extensie van Archimate. Met de modelleertool uit de onderzoeksorganisatie konden oplossingen in het motivatiemodel gerelateerd worden aan aanpassingen in het procesmodel.

In het onderzoek heeft slechts één iteratie plaatsgevonden. In de klankbordbijeenkomsten en met behulp van een enquête achteraf is DQPM getoetst en beproefd. Er zijn een aantal verbeteringen voorgesteld en aantal punten om uit te zoeken in een volgende iteratie. In het kader van dit onderzoek heeft dat niet meer plaats kunnen vinden. Echter, kijkend naar de aard van de veranderingen voor een volgende iteratie, kan wel gesteld worden dat het doel van dit onderzoek gehaald is, de leden van de klankbordgroep hebben aangegeven dat de methode duidelijk en begrijpelijk is en handvatten biedt om problemen met datakwaliteit op te lossen. Door de opzet van de methode is het zowel toe te passen op bestaande als nieuwe processen.

### 7.3. Aanbevelingen voor de praktijk

Er zijn verschillende situaties waarin het voor organisaties zinvol is om deze methode toe te passen. De aanleiding voor toepassing kan voortvloeien uit een auditresultaat, problemen in de operatie, verstoringen in een proces, een belangrijke wijziging in het proces of als onderdeel van een groter kwaliteitsprogramma. In alle gevallen krijgt de organisatie een beter beeld van hoe het gesteld is met de kwaliteit van gegevens in het proces en hoe het de kwaliteit verbeterd kan worden als dat nodig is. In de volgende paragraaf staan nog wel een aantal zaken genoemd voor verder onderzoek, maar ook deze versie van DQPM is genoeg wetenschappelijk onderbouwd om inzicht te kunnen geven.

Voor de onderzoeksorganisatie is de aanbeveling om de methode in een pilotsetting nog een aantal keren te beproeven, de punten voor verder onderzoek mee te nemen en waar nodig de methode te verbeteren. Daarna zou DQPM opgenomen kunnen worden in de verzameling methoden, technieken, hulpmiddelen en voorschriften. Het draagt dan bij aan een van de architectuurprincipes die in de Referentiearchitectuur Gegevenshuishouding opgenomen staat dat kwaliteit van gegevens bij de bron geborgd moet zijn. Referentiearchitecturen hebben een voorschrijvend karakter binnen de onderzoeksorganisatie. Met DQPM wordt aantoonbaar geborgd dat de kwaliteit in de bron beheerst wordt.

Voor de onderzoeksorganisatie geldt dat de IV organisatie grotendeels werkt volgens het SAFE framework, een methodisch kader om lean en agile te werken (Leffingwell, 2019). Om DQPM te gebruiken zouden medewerkers met een beslissende rol in het portfolio, business owners en product owners een voorlichting moeten krijgen met hoe de methode werkt en wat die oplevert. Daarmee kunnen ze de afweging maken ten opzichte van andere portfolio-items.

## 7.4. Aanbevelingen voor verder onderzoek

In de week van het schrijven van dit hoofdstuk deed zich toevallig een incident voor in de onderzoeksorganisatie in relatie tot het proces dat in de case gebruikt is. Voor het inwinnen van een bepaalde stroom waren adresgegevens in het bericht nooit verplicht, maar werden altijd wel gevuld door de aanlevers. Nu was er een nieuwe aanlevers die de adresgegevens niet gevuld had waardoor er in het proces, bij het verifiëren en identificeren van persoonsgegevens, uitval ontstond in zodanig ruime mate dat er sprake was van een productieverstoring. Op zich werkte het proces als ontworpen, gegevens waren niet verplicht, het proces genereerde uitval, maar door de omvang werd het een productieverstoring. Had toepassing van de methode deze verstoring kunnen voorspellen of kunnen laten zien?

Door deze vraag te stellen, valt een van de opmerkingen van de klankbordgroep nu beter op zijn plek. Onder andere in TG.8 en FG.13 (zie bijlage 7) wordt aangegeven om per gevals categorie te onderzoeken wat de methode als resultaat oplevert. Kijkend naar de verstoring van afgelopen week in de organisatie is gevals categorie niet fijnmazig genoeg. Binnen een gevals categorie, bijvoorbeeld een ingewonnen verzekeringsbericht, zijn nog verschillende situaties te onderkennen: een foutief aangeleverd bericht, een bericht zonder verplichte rubrieken, een verkeerd geadresseerd bericht, een bericht dat onwaarschijnlijke bedragen bevat etc. Door het proces te simuleren voor deze verschillende situaties zou een dergelijk probleem voorspeld kunnen worden en zouden aanvullende maatregelen ontworpen kunnen worden.

Dit zou in een volgende iteratie als een verdieping van stap 4 opgenomen kunnen worden, simuleren van het proces voor verschillende situaties in verschillende gevals categorieën. Dit voorbeeld toont aan dat er enerzijds een stap gezet is in het opvullen van tekortkomingen die in de literatuur geconstateerd zijn, maar dat er anderzijds nog verder onderzoek plaats moet vinden om de methode verder te ontwikkelen.

Andere punten die uit het onderzoek naar voren zijn gekomen om mee te nemen in een vervolgonderzoek zijn:

- De aansluiting tussen verbetervoorstellen uit deze methode en de IV voortbrenging, dit kan nog verder doorontwikkeld worden. Daarmee kunnen aangereikte handvatten voor verbetering beter gerealiseerd worden.
- Op activiteitsniveau zijn de gegevensbewerkingen pas zichtbaar en hoe brongegevens getransformeerd worden tot eindproduct. In dit onderzoek werd de transformatie niet als duidelijk beoordeeld. In een vervolgonderzoek dient het activiteitsniveau meegenomen te worden en dient nagegaan te worden of de transformatie daarmee duidelijker gemaakt kan worden.
- Een ander soort proces onderzoeken in de volgende iteratie, bijvoorbeeld een proces waarbij een aangifte getransformeerd wordt tot een opgelegde aanslag. Hiermee kan de kwaliteit van het informatieproduct beter beoordeeld worden.
- Toetsen of de doorgevoerde verbeteringen het beoogde effect hebben.

Een ander punt voor onderzoek, als voorgaande punten onderzocht en verwerkt zijn, is om te kijken naar regelgestuurde processen waarbij regels in een rule engine bepalen wat er gebeurt en er geen aaneenschakeling is van processtappen in een vaste volgorde. Er is een bepaalde uitgangssituatie, er worden regels uitgevoerd die horen bij die situatie waarna er een nieuwe uitgangssituatie is waarbij weer andere regels uitgevoerd kunnen worden. Dit is een andere situatie met andere karakteristieken en andere modellen waarbij nieuw onderzoek nodig is of de methode toepasbaar is.

## 7.5. Reflectie

In het plan van aanpak, hoofdstuk 3, zijn de twee belangrijkste methodologische issues beschreven voor dit type onderzoek: 'rigor' en 'relevance'. Op de relevantie is in het literatuuronderzoek al uitvoerig ingegaan. Er zijn in de literatuur methoden te vinden die herontwerp van processen als uitgangspunt nemen om datakwaliteit te verbeteren. Over het algemeen is er echter nog weinig onderzoek naar gedaan naar het beschrijven en verbeteren van datakwaliteit in de context van BPM (Ofner et al., 2012). Dit onderzoek draagt bij aan verder inzicht in dit onderwerp.

Rigor wordt bereikt door gebruik te maken van bestaande methoden en inzichten in zowel de constructie als de evaluatie van het product. Dat is gedaan door de fasering die in het literatuuronderzoek beschreven is te gebruiken als basis voor DQPM. De inventarisatie van Batini et al. (2009) geeft aan dat de meeste methoden om datakwaliteit te verbeteren in een procesaanpak drie fases bevatten: beschrijvend, assessment en verbetering. Die vormen de basis. De tekortkomingen die gevonden zijn in de bestaande methoden zijn zoveel mogelijk weg genomen in de ontwikkeling van de stappen van DQPM. Verder is DQPM gebaseerd op inzichten uit de literatuur dat een informatieproduct geproduceerd wordt analoog aan de werkwijze in de productie-industrie. Bij de evaluatie van DQPM is gebruik gemaakt van criteria die in de literatuur zijn gevonden. Deze zijn vervolgens ook gewogen volgens Analytical Hierarchy Process, AHP. De weging is gedaan door de leden van de klankbordgroep. Dat is gedaan omdat zij twee aspecten vertegenwoordigen uit het probleemgebied (Environment) uit het Information Systems Research Framework van Hevner et al. (2004): mensen en organisaties. Daarmee wordt ook de relevantie versterkt, de leden van de klankbordgroep geven de belangrijkste eigenschappen van het probleemgebied aan.

De pragmatische validiteit die door van Aken et al. (2016) genoemd wordt, uitgebreid testen in het werkveld, is in dit onderzoek beperkt in verband met de beschikbare tijd. Wel is de aanpak en de uitvoering beschreven zodat het onderzoek in andere organisaties en andere contexten op dezelfde manier uitgevoerd kan worden.



## Referenties

- Aljumaili, M., Karim, R., & Tretten, P. (2016). Metadata-based data quality assessment. *VINE Journal of Information and Knowledge Management Systems*, 46(2), 232–250.
- Başkarada, S., & Koronios, A. (2014). A Critical Success Factor Framework for Information Quality Management. *Information Systems Management*, 31(4), 276–295.
- Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Comput. Surv.*, 41(3), 1–52.
- Besselink, C. (2010, April). Gegevenskwaliteit in overheidsnetwerken. *Informatie*, 209–221.
- Cao, L., & Zhu, H. (2013). Normal Accidents: Data Quality Problems in ERP-Enabled Manufacturing. *Journal of Data and Information Quality*, 4(3).
- Cappiello, C., Caro, A., Rodriguez, A., Caballero, I., Caro, A., & Rodríguez, A. (2013). An Approach To Design Business Processes Addressing Data Quality Issues. In *ECIS 2013 Completed Research* (p. Paper 216).
- Glowalla, P., & Sunyaev, A. (2014). Process-driven Data Quality Management: A Critical Review on the Application of Process Modeling Languages. *J. Data and Information Quality*, 5(1–2), 7:1–7:30.
- Goepel, K. D. (2013). No Title. In *Implementing the Analytic Hierarchy Process as a Standard Method for Multi-Criteria Decision Making In Corporate Enterprises – A New AHP Excel Template with Multiple Inputs, Proceedings of the International Symposium on the Analytic Hierarchy Process*, Kua (pp. 1–10).
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *Two Paradigms Characterize Much of the Research in the Information Systems Discipline: Behavioral Science and Design Science. The Behavioral- Science Paradigm Seeks to Develop and Verify Theories That Explain or Predict Human or Organi- Zational Behavior.*, 28(1), 75–105.
- Lee, Y., Pipino, L., Funk, J., & Wang, R. (2006). *Journey to data quality. Computer* (Vol. 1).
- Lee, Y. W., Pipino, L. L., Funk, J. D., & Wang, R. Y. (2009). *Journey to Data Quality*. The MIT Press.
- Lee, Y. W., & Strong, D. M. (2003). Knowing-Why About Data Processes and Data Quality. *Journal of Management Information Systems*, 20(3), 13–39.
- Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: A methodology for information quality assessment. *Information and Management*, 40(2), 133–146.
- Leffingwell, D. (2019). Scaled Agile Framework.
- Madnick, S., & Hongwei Zhu. (2006). Improving data quality through effective use of data semantics. *Data & Knowledge Engineering*, 59(2), 460–475.
- McGilvray, D. (2008). *Data Quality Projects: Ten Steps to Quality Data and Trusted Information*. Morgan Kaufmann.
- Mezzanzanica, M., Boselli, R., Cesarini, M., & Mercorio, F. (2015). A model-based evaluation of data quality activities in KDD. *Information Processing and Management*, 51(2).

- Michael Mielke. (n.d.). Information Quality and BI a strategic alliance.
- Mielke, M. (2005). IQ principles in software development. In *Proceedings of the International Conference on Information Quality*.
- Moossavizadeh, S. M. H., Mohsenzadeh, M., & Arshadi, N. (2012). A new approach to measure believability dimension of data quality. *Management Science Letters*, 2(7), 2565–2570.
- Nass, E., & Scheibmayer, M. (2011). Defining a research framework for the business impact of data management. In K.-D. Thoben, V. Stich, & A. Imtiaz (Eds.), *Proceedings of the 2011 17th International Conference on Concurrent Enterprising* (pp. 1–8). Aachen, Germany.
- Obers, G. J., & Achterberg, K. (2014). *Grip op processen in organisaties* (Tweede dru). Zaltbommel: Van Haren Publishing.
- Object Management Group. (2014). Documents associated with BPMN version 2.0.2.
- Object Management Group (OMG). (2011). Business Process Model and Notation (BPMN) Version 2.0. *Business*, 50(January), 170.
- Object Management Group (OMG). (2013). Archimate 2.1 specification.
- Ofner, M. H., Otto, B., & Österle, H. (2012). Integrating a data quality perspective into business process management. *Business Process Management Journal*, 18(6), 1036–1067.
- Otten, S., & Karels, A. (2018). *Referentiearchitectuur Gegevenshuishouding*.
- Panwar, M. (2015). Application of Six-Sigma for Data Quality Improvement in an Insurance Company, 4(09).
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77.
- Petkov, P., & Helfert, M. (2013). A methodology for analyzing and measuring semantic data quality in service oriented architectures. *Proceedings of the 14th International Conference on Computer Systems and Technologies - CompSysTech '13*, 201–208.
- Pipino, L. L., Lee, Y. W., Wang, R. Y., Lowell Yang Lee, M. W., & Yang, R. Y. (2002). Data Quality Assessment. *COMMUNICATIONS OF THE ACM April*, 45(4ve).
- Rodríguez, A., Caro, A., Cappiello, C., & Caballero, I. (2012). A BPMN Extension for Including Data Quality Requirements in Business Process Modeling BT - Business Process Model and Notation, 116–125.
- Roser, C., Lorentzen, K., & Deuse, J. (2014). Reliable shop floor Bottleneck detection for flow lines through process and inventory observations. *Procedia CIRP*, 19(C), 63–68.
- Saaty, T. L. (1980). *The Analytic Hierarchy Process*. McGraw-Hill. New York.
- Saunders, M. N. K., Lewis, P., & Thornhill, A. (2012). *Research methods for business students*. Harlow, England; New York: Pearson.
- Shankaranarayanan, G., Wang, R. Y., & Ziad, M. (2000). IP-MAP: Representing the Manufacture of an Information Product. *Proceedings of the 2000 Conference on*

*Information Quality*, 1–16.

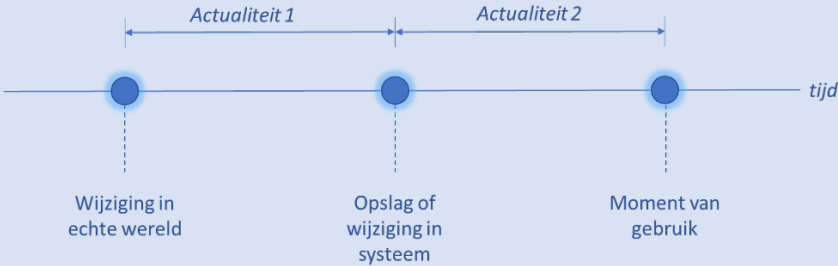
- Storey, V. C., Dewan, R. M., & Freimer, M. (2012). Data quality: Setting organizational policies. *Decision Support Systems*, 54(1), 434–442.
- Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data Quality in Context. *Commun. ACM*, 40(5), 103–110.
- Todoran, I.-G., Lecornu, L., Khenchaf, A., & Caillec, J.-M. Le. (2015). A Methodology to Evaluate Important Dimensions of Information Quality in Systems. *J. Data and Information Quality*, 6(2–3), 11:1–11:23.
- van Aken, J., Chandrasekaran, A., & Halman, J. (2016). Conducting and publishing design science research: Inaugural essay of the design science department of the Journal of Operations Management. *Journal of Operations Management*, 47–48, 1–8.
- Van der Aalst, W. M. P. (2013). Business Process Management: A Comprehensive Survey. *ISRN Software Engineering*, 2013, 1–37.
- Vieveen, P., Geerlings, H., Nederpelt van, P., Kolthof, H., Rubrech, J., Smit, P., & Wijsman, T. (2016). Code voor Informatiekwaliteit 2016.
- Wahab, A. N. A., Mukhtar, M., & Sulaiman, R. (2013). A Conceptual Model of Lean Manufacturing Dimensions. *Procedia Technology*, 11, 1292–1298.
- Wang, R., Reddy, M., & Kon, H. (1995). Toward quality data: An attribute-based approach. *Decision Support Systems*, 13(3–4), 349–372.
- Wee, H. M., & Wu, S. (2009). Lean supply chain and its effect on product cost and quality: a case study on Ford Motor Company. *Supply Chain Management: An International Journal*, 14(5), 335–341.
- Wokke, P. (2018). *BPMN conventiedocument Belastingdienst*. Apeldoorn.
- Yang, L., Neagu, D., Cronin, M. T. D., Hewitt, M., Enoch, S. J., Madden, J. C., & Przybylak, K. (2013). Towards a fuzzy expert system on toxicological data quality assessment. *Molecular Informatics*, 32(1), 65–78.
- Zou, X., Settini, R., & Cleland-Huang, J. (2010). Improving automated requirements trace retrieval: A study of term-based enhancement methods. *Empirical Software Engineering*, 15(2), 119–146.

## Bijlage 1 Dimensies van datakwaliteit

Tabel 18 Kwaliteitsdimensies

| Invalshoek            | Dimensie          | Definitie   | Metten  |
|-----------------------|-------------------|---|---|
| Intrinsieke kwaliteit | Juistheid         | (Batini et al., 2009) maken onderscheid tussen syntactische juistheid en semantische juistheid. Semantische juistheid geeft aan of een waarde w een afspiegeling vormt van de werkelijkheid. Methoden voor datakwaliteit kijken alleen naar syntactische juistheid. Dat is de mate waarin een waarde w de elementen van een overeenkomstig definitiedomein D benadert.  | Y. Lee, Pipino, Funk, & Wang (2006) geven de volgende formule om juistheid te berekenen:<br>$\text{Juistheid} = 1 - \frac{\text{Aantal syntactisch onjuiste dataelementen}}{\text{Totaal aantal dataelementen}}$  |
|                       | Geloofwaardigheid | (Moossavizadeh et al., 2012) geven aan dat geloofwaardigheid een moeilijk te definiëren en te meten begrip is omdat het voor meerdere uitleg vatbaar is. Zij definiëren geloofwaardigheid als de mate waarin data geaccepteerd als waar of als item dat waar, echt en geloofwaardig lijkt in een specifieke omgeving en in overeenstemming met relevante regels.<br>Geloofwaardige data in een systeem kan of <u>echt</u> waar zijn en consistent met actuele data, of waar, echt en geloofwaardig <u>lijken</u> te zijn. | $\text{Geloofwaardigheid (DI)} = \frac{(W_J \cdot J) + (W_O \cdot O) + (W_R \cdot R) + (W_X \cdot X)}{\sum_i W_i}$<br>; $i \in \{J, O, R, X\}$<br>Waarbij:<br>J – Juistheid, de mate waarin data in een systeem overeenstemt met de afgebeelde werkelijkheid. Hoewel een verband tussen juistheid en geloofwaardigheid aanwezig is, kan data juist zijn en toch ongeloofwaardig.<br>O – Consistentie met rationele en organisatorische regels, uitgedrukt in een getal tussen 0 en 1, waarbij 0 inconsistent is en 1 consistent. Kennis van experts is nodig om de regels op te stellen die de mate van consistentie bepalen.<br>R – Bruikbaarheid van resources. Hoe bruikbaarder de resource, hoe geloofwaardig de data. Bruikbaarheid is het product van de betrouwbaarheid van de resource en de relevantie van de resource voor de data.<br>X – Consistentie met eerdere waarnemingen, uitgedrukt in een getal tussen 0 en 1, waarbij 0 inconsistent is en 1 consistent. Ook hier is kennis van experts nodig om de regels op te stellen die de mate van consistentie bepalen.<br><br>$W_J$ =Gewicht juistheid, $W_O$ = Gewicht consistentie met rationele en organisatorische regels, $W_R$ = Gewicht bruikbaarheid resources en $W_X$ = Gewicht consistentie met eerdere waarnemingen. |
|                       | Consistentie      | (Batini et al., 2009) omschrijft consistentie als de mate waarin semantische regels worden  | Y. Lee et al. (2006) geven de volgende formule:   |

| Invalshoek | Dimensie   | Definitie  | Meten  |
|------------|------------|--|--|
|            |            | overschreden die van toepassing zijn op de verzameling data items. Een belangrijke groep semantische regels zijn integriteitsregels. Deze bevatten op hun beurt intra-relatieregels en inter-relatieregels. Intra-relatieregels geven het waardebereik voor een dataelement aan, bijvoorbeeld een leeftijd die tussen 0 en 125 mag zijn of als iemand pensioen geniet, mag de leeftijd niet lager zijn dan 62. In termen van een relationele database hebben intra-relatieregels betrekking op attributen binnen één tabel (Besselink, 2010). Interrelatieregels hebben betrekking op attributen van verschillende relaties of in termen van relationele databases op attributen uit verschillende tabellen. | $1 - \frac{\text{Aantal instanties dat semantische regel overschrijdt}}{\text{Aantal uitgevoerde controles op die regel}}$ <p>Bij meerdere regels wordt het aantal gesommeerd in zowel de teller als de noemer gesommeerd.</p>   |
| Context    | Relevantie | (Todoran, Lecornu, Khenchaf, & Caillec, 2015) omschrijft relevantie als de mate waarin gegevens toepasbaar zijn voor en bruikbaar zijn voor een specifieke taak.   | <p>Pipino et al. (2002) beschrijven dat dimensies objectief en subjectief vastgesteld kunnen worden. Deze dimensie is alleen subjectief vast te stellen door middel van een 'user assessment' (Todoran et al., 2015). (Y. W. Lee &amp; Strong, 2003) scherpen dit aan tot data consumers, zij hebben de kennis om deze dimensie te beoordelen. (Michael Mielke, n.d.) geeft de volgende formule:</p> $\text{Relevantie} = \frac{\text{Aantal geleverde gegevens die relevant zijn voor afnemer}}{\text{Totaal aantal geleverde gegevens}}$ |
|            | Tijdigheid | <p>Batini et al. (2009) geeft in zijn inventarisatie van deze dimensie weer dat er drie aspecten zijn van tijdigheid:</p> <ul style="list-style-type: none"> <li>– Actualiteit, geeft aan hoe oud de informatie is ten opzichte van het moment dat het opgeslagen of gewijzigd werd. Van belang voor gegevens die snel verouderen.</li> <li>– Tijdigheid, de mate waarin tijdig over gegevens beschikt kan worden.</li> </ul>  | <p>Batini et al. (2009) geeft verschillende mogelijkheden om te meten. Hier geven we de metrieken die het meest aansluiten bij de gegeven definities. Onderstaande afbeelding geeft de twee definities van actualiteit weer die in dit onderzoek gebruikt worden, de tijd tussen de verschillende tijdstipmomenten.</p>  |

| Invalshoek | Dimensie     | Definitie   | Metten  |
|------------|--------------|---|---|
|            |              | <ul style="list-style-type: none"> <li>Volatiliteit, de (in)stabiliteit van informatie, de frequentie van wijzigingen in een waarde van een attribuut (Bovee et al. 2001) of de periode waarin gegevens overeenkomen met de waarde in de 'echte wereld' (Jarke et al. 1995).</li> </ul> <p>In de praktijk worden de verschillende aspecten door elkaar gebruikt, terwijl ze hetzelfde beogen te beschrijven. Soms liggen ze ook dicht tegen elkaar aan. Een van de definities uit het overzicht van Batini is bijvoorbeeld de tijd tussen een wijziging in de echte wereld en het moment waarop de wijziging doorgevoerd is in het systeem. Is hier sprake van actualiteit of tijdigheid?</p>   |  <p><i>Figuur 21 Definities actualiteit</i></p> <p><i>Tijdigheid = % gegevens waarover tijdig beschikt kan worden t. b. v. uitvoering van processen</i></p> <p><i>Volatiliteit = Duur waarin data geldig is</i></p>  |
|            | Volledigheid | <p>(Batini et al., 2009) geeft een overzicht van de verschillende definities in de literatuur van volledigheid. Ze variëren vooral in de context waarin volledigheid beoordeeld wordt; informatiesystemen, datawarehouses of entiteiten. Volledigheid kan gezien worden als de mate waarin de dataelementen de totale relevante populatie vertegenwoordigen (dekkingsgraad). Volledigheid kan ook gezien worden als de mate waarin alle velden in een regel, kolom of tabel gevuld zijn met een relevante waarde (vullingsgraad). In beide gevallen is belangrijk om te weten waarom een waarde niet gevuld is:</p> <ul style="list-style-type: none"> <li>Waarde bestaat niet (iemand heeft geen mobiel nummer)</li> <li>Waarde bestaat wel, maar is niet bekend (onbekend of iemand een mobiel nummer heeft)</li> <li>Onbekend of waarde bestaat</li> </ul> | <p>De vullingsgraad kan volgens Batini et al. (2009) als volgt berekend worden, waarbij de boolean waardes gebruikt worden om te bepalen of een veld compleet is of niet.</p> $Vullingsgraad = \frac{Aantal\ complete\ velden}{Totaal\ aantal\ velden}$ <p>Voor de dekkingsgraad verwijst Batini naar Redman die de volgende formule geeft:</p> $Dekkingsgraad = \frac{Aantal\ objecten\ in\ gegevensverzameling}{Aantal\ objecten\ in\ de\ populatie}$ |

| Invalshoek           | Dimensie                 | Definitie  | Metten   |
|----------------------|--------------------------|--|--|
|                      |                          | (Batini et al., 2009) geeft aan dat ten behoeve van een data quality assessment ieder attribuut een boolean waarde toegevoegd moet krijgen die een indicatie bevat of het compleet is of niet.   |  |
| <b>Representatie</b> | Interpreteerbaarheid     | Wang et al. (1995) geeft als definitie de mate waarin de afnemer van data de semantiek en de syntax van de data begrijpt. Het aspect semantiek wordt in de volgende dimensie verder toegelicht.  | Batini et al. (2009) geven een formule om interpreteerbaarheid te kunnen berekenen, het aantal rijen met interpreteerbare data, dat is, sleutelwaarden waarvoor documentatie beschikbaar is. Wang et al. (1995) geven aan dat een dimensie is die gemeten kan worden door een user survey. |
|                      | Semantische consistentie | <p>Petkov &amp; Helfert (2013) stellen dat semantiek als dimensie voor datakwaliteit in veel methoden ontbreekt. De semantiek van data kan in de loop van de tijd veranderen (Madnick &amp; Hongwei Zhu, 2006). Daardoor kan de representatieve of de ontologische semantiek van een bron of een ontvanger zich ontwikkelen. Verder speelt het tijdsaspect mee. Madnick &amp; Hongwei Zhu (2006) komen tot de volgende categorisering van heterogeniteit die van invloed is op datakwaliteit.</p> <p><u>Representatieve heterogeniteit</u>: hetzelfde concept heeft verschillende representaties in verschillende bronnen of afnemers. Bijvoorbeeld een bepaalde dag kan weergegeven worden als '21 maart 2017', '21-03-17' of '03-21-2017'.</p> <p><u>Tijdelijke representatieve heterogeniteit</u>: de representatie binnen een bepaalde bron of afnemer verschilt. Bijvoorbeeld een artikeltabel met prijzen die eerst in guldens weergegeven werden en later in euro's.</p> <p><u>Ontologische heterogeniteit</u>: dezelfde term heeft in verschillende contexten een andere betekenis. Bijvoorbeeld het begrip 'inkomen' heeft op een belastingformulier een andere betekenis dan op de salarisstrook van een werknemer.</p> <p><u>Tijdelijke ontologische heterogeniteit</u>: de betekenis heeft binnen een bepaalde bron of afnemer in de</p> | <p>Batini et al. (2009) geven een algemene manier om consistentie te meten die ook bruikbaar is voor semantische consistentie.</p> $\text{Semantische consistentie} = \frac{\text{Aantal consistente waarden}}{\text{Totaal aantal waarden}}$  |

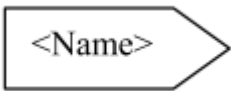

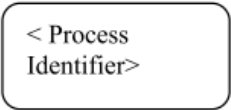

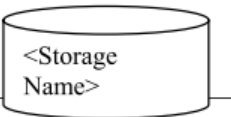



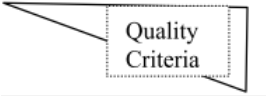
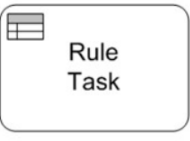

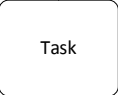
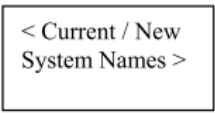

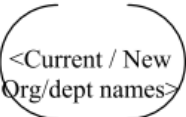



| Invalshoek              | Dimensie         | Definitie   | Metten  |
|-------------------------|------------------|---|---|
|                         |                  | loop van de tijd een andere betekenis gekregen. Het begrip nettowinst bijvoorbeeld kan een andere betekenis krijgen in dezelfde database als de regelgeving wijzigt die bepaalt wat er onder nettowinst wordt verstaan.   |   |
|                         | Begrijpelijkheid | Besselink (2010) definieert begrijpelijkheid als de mate van beschikbaarheid van documentatie en metadata om het bericht correct te kunnen interpreteren. Daarmee zit deze dimensie dicht tegen interpreteerbaarheid aan. | Todoran et al. (2015) geven aan dat deze dimensie het beste met een user survey gemeten kan worden.   |
| <b>Toegankelijkheid</b> | Toegankelijkheid | Toegankelijkheid is indicatie van het gemak waarmee de gebruiker over data kan beschikken (Batini et al., 2009).  | <p>Batini et al. (2009) geven de volgende formule voor het berekenen van toegankelijkheid waarin het tijdsinterval van vraag tot beschikbaarheid van gegevens een rol speelt en het tijdsinterval totdat beschikbaar stellen van gegevens geen rol meer speelt.</p> $Toegankelijkheid = \max(0; 1 - \frac{(Tijdstip\ levering - Tijdstip\ verzoek)}{(Tijdstip\ deadline - Tijdstip\ verzoek)})$ <p>Daarnaast kan ook een user survey gebruikt worden om deze dimensie te meten.</p> |
|                         | Beveiliging      | Beveiliging geeft aan in welke mate de gegevens beschermd zijn tegen ongeautoriseerde toegang (Y. W. Lee et al., 2002).   | De metriek die Batini et al. (2009) geven is het aantal 'zwakke' loginpogingen. Verder kan ook een user survey gebruikt worden om deze dimensie te meten.   |



## Bijlage 2 Concepten IP-MAP vs BPMN concepten

Onderstaande tabel geeft de modelleerconcepten van IP-MAP aan zoals genoemd in het artikel van Shankaranarayanan et al. (2000) en het concept uit BPMN (Object Management Group, 2014) dat daar het meest bij aansluit.

Tabel 19 IP Map concepten vs BPMN concepten

| Concept IP Map  | Omschrijving                                    | Concept BPMN  | Omschrijving                  |
|---|---|---|-------------------------------|
|    | Data source, data vendor, point of origin       |    | Start event                   |
|    | Process   |    | Task of Sub process           |
|    | Data/information storage                        |    | Data store                    |
|   | Decision  |   | Gateway                       |
|  | Quality, evaluation, check                      | <br><br> | Rule task of Lezen data store |
|  | Information system boundary                     |   | Other domain of swim lane     |
|  | Organizational boundary                         |   | Lane of pool                  |
|  | Data sink, consumer block, point of destination |   | Black box pool                |

## Bijlage 3 Requirements aan procesmodellen en datakwaliteit

Shankaranarayanan et al. (2000) noemen in de door hun ontwikkelde methode IP-MAP de volgende requirements waar een datakwaliteitsgeoriënteerd procesmodel aan moet voldoen.

1. Methode moet op een systematische manier de processen laten zien die nodig zijn om een informatieproduct voort te brengen.
2. De belangrijkste fasen in het proces en kritieke fasen die datakwaliteit beïnvloeden moeten gevisualiseerd kunnen worden.
3. Het model moet inzicht geven in bottlenecks in het proces.
4. Het model moet inzicht geven in de tijd die nodig is om een informatieproduct voort te brengen.
5. Het model moet duidelijk maken wie eigenaar is van de processen die informatieproducten voortbrengen.
6. Het model moet ondersteunend zijn in het implementeren van kwaliteitsmaatregelen.
7. Het model moet inzicht geven in de informatiesysteemgrenzen van de systemen die nodig zijn in de verschillende processen om een informatieproduct voort te brengen.
8. Het model moet het mogelijk maken om de kwaliteit van het informatieproduct meetbaar te maken aan de hand van de relevante kwaliteitsdimensies.

Ofner et al. (2012) komen in hun onderzoek tot de volgende requirements.

1. De methode moet concepten en procedures bevatten om datakwaliteit in procesontwerpen te beoordelen in projecten die gericht zijn op herontwerp van bedrijfsprocessen.
2. De concepten van de methode moeten zoveel mogelijk gebaseerd zijn op standaard BPM-methoden.
3. De methode moet het beoordelen van datakwaliteit op verschillende abstractieniveaus mogelijk maken ten behoeve van verschillende stakeholders (hoog niveau voor managers, gedetailleerd niveau voor ontwikkelaars).
4. De methode moet concepten bevatten om het belang van de business impact en waarde van datakwaliteit te kunnen laten zien.
5. De opgeleverde procesmodellen moeten bruikbaar zijn als input voor het systeemontwikkelingstraject.
6. De methode moet rekening houden met de relevante context van data zodat een gedifferentieerde representatie van data mogelijk is.

Glowalla & Sunyaev (2014) hebben eisen verzameld ten aanzien van representatie. Ze stellen daarbij dat verder onderzoek nodig is. Hun basis wordt gevormd door de eisen die gesteld worden aan de IP-Map en daaromheen hebben ze aanvullende eisen gezet. Onderstaande afbeelding laat e.e.a. zien.

|                |  |  |  |   |  |
|----------------|--|--|--|---|--|
| Complementary: |  | <ul style="list-style-type: none"> <li>transformation process</li> </ul> | <ul style="list-style-type: none"> <li>swim lane</li> </ul>          | <ul style="list-style-type: none"> <li>time axis</li> <li>control flow</li> </ul> | <ul style="list-style-type: none"> <li>quality dimensions</li> <li>quality metrics</li> <li>additional data constructs (e.g. reports)</li> </ul> |
| IP-MAP:        | Data Source<br>Data Sink<br>Data Storage | Process<br>Quality check<br>Decision                                     | IS-Boundary<br>Organizational Boundary<br>IS-Organizational Boundary | Raw Data flow<br>Component Data flow<br>IP flow                                   |  |
| Redundant:     |  |  |  |   |  |

Figuur 22 Representatie eisen PDDQM (Glowalla & Sunyaev, 2014)

Naast deze inhoudelijke eisen stellen zij dat de gebruikte modellen vertrouwd moeten zijn voor stakeholders om ervoor te zorgen dat ze ook daadwerkelijk gebruikt gaan worden. Verder moeten modellen begrijpelijk zijn en niet te complex. Wat dat betreft is het balanceren tussen begrijpelijkheid voor gebruikers en het beoogde doel van het model.

## Bijlage 4 Beslistabellen stap 4 DQPM

De eerste beslissing, de mate waarin requirements uitgewerkt zijn in het procesmodel en de documentatie ziet er in een tabel als volgt uit.

Tabel 20 Beslissing uitwerking requirements in model en documentatie

| Uitwerking requirements in model en documentatie |                           |               |                            |                                    |                                  |              |  |
|--|---------------------------|---------------|----------------------------|------------------------------------|----------------------------------|--------------|--|
| U  | Aanwezigheid documentatie |               | Uitwerking requirements DQ |                                    | Volledigheid kwaliteitsdimensies |              | Uitwerking requirements in model en documentatie |
| 1  | is                        | Niet aanwezig | is                         | Uitgewerkt of afleidbaar           | is                               | Volledig     | Goed   |
| 2  | is                        | Niet aanwezig | is                         | Uitgewerkt of afleidbaar           | is                               | Gedeeltelijk | Redelijk   |
| 3  | is                        | Niet aanwezig | is                         | Niet uitgewerkt en niet afleidbaar |                                  | –            | Slecht   |
| 4  | is                        | Aanwezig      | is                         | Niet uitgewerkt en niet afleidbaar |                                  | –            | Slecht   |
| 5  | is                        | Aanwezig      | is                         | Uitgewerkt of afleidbaar           | is                               | Volledig     | Goed   |
| 6  | is                        | Aanwezig      | is                         | Uitgewerkt of afleidbaar           | is                               | Gedeeltelijk | Redelijk   |

De tweede deelbeslissing, de mate waarin datakwaliteit in uitvoering van het proces vastgesteld wordt, staat in de volgende beslistabel uitgewerkt.

Tabel 21 Beslissing datakwaliteit in procesuitvoering

| <b>Datakwaliteit in procesuitvoering</b> |                                |                |                         |                |  |
|--|--------------------------------|----------------|-------------------------|----------------|--|
| <b>U</b>                                 | <b>Interviews stakeholders</b> |                | <b>Kwaliteitsmeting</b> |                | <b>Datakwaliteit in procesuitvoering</b> |
| 1  | is                             | Positief beeld | is                      | Positief beeld | Goed                                     |
| 2  | is                             | Positief beeld | is                      | Negatief beeld | Redelijk                                 |
| 3  | is                             | Positief beeld | is                      | Neutraal beeld | Redelijk                                 |
| 4  | is                             | Negatief beeld | is                      | Positief beeld | Redelijk                                 |
| 5  | is                             | Negatief beeld | is                      | Negatief beeld | Slecht                                   |
| 6  | is                             | Negatief beeld | is                      | Neutraal beeld | Slecht                                   |
| 7  | is                             | Neutraal beeld | is                      | Positief beeld | Redelijk                                 |
| 8  | is                             | Neutraal beeld | is                      | Negatief beeld | Slecht                                   |
| 9  | is                             | Neutraal beeld | is                      | Neutraal beeld | Redelijk                                 |

De uitkomst van de twee deelbeslissingen leiden samen tot een oordeel over de score van kwaliteitsdimensies voor een BPMN concept. Onderstaande tabel geeft de afleiding van de score weer.

Tabel 22 Beslissing score kwaliteitsdimensie voor BPMN concept

| Score kwaliteitsdimensie voor BPMN concept |                                   |          |  |          |  |   |
|--|-----------------------------------|----------|--|----------|--|---|
| U  | Datakwaliteit in procesuitvoering |          | Uitwerking requirements in model en documentatie |          | Score kwaliteitsdimensie voor BPMN concept | Annotatie   |
| 1  | is                                | Goed     | is   | Goed     | Goed                                       |   |
| 2  | is                                | Goed     | is   | Redelijk | Goed                                       | Documentatie is dan mogelijk geen goede afspiegeling van de werkelijkheid |
| 3  | is                                | Goed     | is   | Slecht   | Redelijk                                   | Documentatie is dan mogelijk geen goede afspiegeling van de werkelijkheid |
| 4  | is                                | Redelijk | is   | Goed     | Redelijk                                   |   |
| 5  | is                                | Redelijk | is   | Redelijk | Redelijk                                   |   |
| 6  | is                                | Redelijk | is   | Slecht   | Slecht                                     |   |
| 7  | is                                | Slecht   | is   | Goed     | Redelijk                                   | Model iken voor een definitief oordeel                                    |
| 8  | is                                | Slecht   | is   | Redelijk | Slecht                                     |   |
| 9  | is                                | Slecht   | is   | Slecht   | Slecht                                     |   |

## Bijlage 5 Verantwoording samenstelling criteria

De criteria worden gebruikt om het model op te toetsen. Het onderzoek wordt uitgevoerd bij de Belastingdienst waar de modelleertaal Business Process Model Notation (BPMN) gebruikt wordt, een standaard van de Object Management Group (OMG). Om de klankbordgroep de criteria te laten wegen, wordt Analytic Hierarchy Process (AHP) toegepast, een methode die het nemen van beslissingen en het wegen van criteria ondersteunt (Saaty, 1980).

Bij AHP worden de criteria paarsgewijs tegen elkaar afgezet en gewogen waarmee een totaalscore gemaakt kan worden. Dit kan samengesteld worden uit de weging en score van meerdere deelnemers. De tooling om AHP te ondersteunen kent een aantal beperkingen. In de modules Business Processes en IT Governance is het tool AH Priest gebruikt, een open source app voor Android of IOS. Deze kent als beperking dat hij binnen de Belastingdienst lastig te gebruiken is omdat op de werkplekken van de Belastingdienst geen andere software geïnstalleerd mag worden. Verder wordt de lijst met koppels van criteria in AH Priest erg lang, er wordt gevraagd om meer dan honderd paren te scoren. De vraag is of dat een betrouwbaar resultaat oplevert. Goepel (2013) heeft een Excel spreadsheet beschikbaar gesteld die eenvoudiger te gebruiken is voor leden van de klankbordgroep. Deze kent een andere beperking, dat er maximaal tien criteria gebruikt kunnen worden.

Dat was de aanleiding om nog eens kritisch te kijken naar de criteria die uit het literatuuronderzoek naar voren zijn gekomen. In de methoden, technieken, hulpmiddelen en voorschriften (MTHV's) die de Belastingdienst hanteert is gekeken welke criteria uit het literatuuronderzoek al geborgd zijn in de MTHV's en niet meer als criterium gebruikt hoeven te worden om te scoren. Onderstaande tabel geeft de criteria uit het literatuuronderzoek weer en geeft aan of en zo ja waar deze in het MTHV's geborgd is. De criteria die meegenomen moeten worden in het ontwerp zijn onderstreept.

Tabel 23 Criteria afgezet tegen standaarden MTHV's

| Soort                 | Criterium   | Geborgd in MTHV's of eis aan het ontwerp  |
|-----------------------|---|---|
| <b>Uitgangspunten</b> | Modellen moeten aansluiten bij standaard BPM-methoden.  | Geborgd in MTHV's<br>De Belastingdienst gebruikt BPMN, een standaardtaal van de OMG. De Belastingdienst heeft geen aanpassingen gedaan op het metamodel van BPMN, wel zijn met behulp van configuratiemogelijkheden in de tool op een aantal plaatsen extra documentatievelden toegevoegd en extra tekstuele eigenschappen. (Wokke, 2018) |
|                       | <u>Modellen moeten vertrouwd zijn voor stakeholders, begrijpelijk zijn en niet te complex.</u>                                | Eis aan het ontwerp.  |
| <b>Algemene eisen</b> | De methode moet op een systematische manier de processen laten zien die nodig zijn om een informatieproduct voort te brengen. | Geborgd in de MTHV's<br>Bij de Belastingdienst worden processen beschreven op het niveau van ketenproces, bedrijfsproces, werkproces, processtap en activiteit. De afbakening van het niveau tot en met werkproces vindt plaats in de architectuur. Keten- en bedrijfsprocessen zijn  |

| Soort                                     | Criterium   | Geborgd in MTHV's of eis aan het ontwerp   |
|---|---|--|
|   |   | gericht op het voortbrengen van producten of diensten. (Wokke, 2018)   |
|   | <u>Het model moet inzicht geven in bottlenecks in het proces.</u>   | Eis aan het ontwerp. Bedoeld wordt bottlenecks in de doorstroming.   |
|   | Het model moet duidelijk maken wie eigenaar is van de processen die informatieproducten voortbrengen.   | Geborgd in de MTHV's<br>In de processen is aangegeven wie responsible, accountable, consulted en informed is (RACI). (Wokke, 2018) |
|   | <u>De opgeleverde procesmodellen moeten bruikbaar zijn als input voor het systeemontwikkelingstraject.</u>  | Eis aan het ontwerp.   |
|   | <u>Het model moet de raw dataflow, component dataflow en IP-flow kunnen laten zien.</u>   | Eis aan het ontwerp.   |
|   | Het model moet de grenzen van betrokken organisaties, onderdelen daarvan en systemen laten zien.  | Geborgd in de MTHV's.  |
|   | <u>Het transformatieproces moet duidelijk zijn.</u>   | Eis aan het ontwerp.   |
| <b>Eisen datakwaliteit in procesmodel</b> | <u>De belangrijkste fasen in het proces en kritieke fasen die datakwaliteit beïnvloeden moeten gevisualiseerd kunnen worden.</u>  | Eis aan het ontwerp.   |
|   | <u>Kwaliteitscontroles zijn expliciet zichtbaar.</u>  | Eis aan het ontwerp.   |
|   | Het model moet inzicht geven in de tijd die nodig is om een informatieproduct voort te brengen.   | Geborgd in de MTHV's.<br>In de eigenschappen van een werkproces kan de doorlooptijd van het proces                                 |
|   | <u>Het model moet ondersteunend zijn in het implementeren van kwaliteitsmaatregelen.</u>  | Eis aan het ontwerp.   |
|   | <u>Het model moet het mogelijk maken om op verschillende abstractieniveaus de kwaliteit van het informatieproduct meetbaar te maken aan de hand van de relevante kwaliteitsdimensies.</u> | Eis aan het ontwerp.   |
|   | <u>De methode moet concepten bevatten om het belang van de business impact en waarde van datakwaliteit te kunnen laten zien.</u>  | Eis aan het ontwerp.   |
|   | <u>De methode moet rekening houden met de relevante context van data zodat een gedifferentieerde representatie van data mogelijk is.</u>  | Eis aan het ontwerp.   |
|   |   |  |
|   |   |  |

Dit levert een beperktere set criteria op en een beter inzicht in welk criterium nieuw is voor de toevoeging van datakwaliteit en welk criterium geborgd is in de huidige standaarden van de



Belastingdienst. Dit betekent overigens wel dat het ontwerp uit dit onderzoek getoetst moet worden op het criterium dat al geborgd is in de standaarden. Ook het nieuwe ontwerp moet (Goepel, 2013) daaraan voldoen.

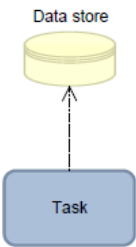
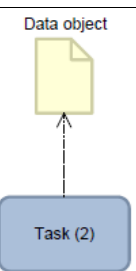
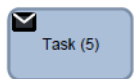
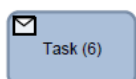
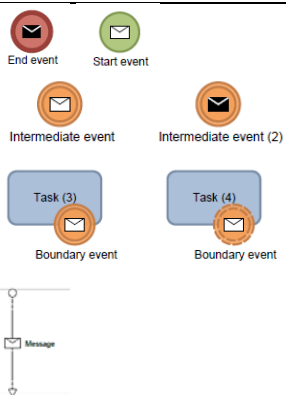
Ook de beperkte set levert meer op dan tien criteria. Op de website van Goepel (2013) is ook een online omgeving te vinden om een AHP analyse mee uit te voeren. Deze kan meer dan tien criteria aan en levert toch een overzichtelijke lijst om in te vullen voor de leden van de klankbordgroep. Na invulling van de klankbordgroep is het resultaat als volgt.

Tabel 24 Weging criteria door klankbordgroep

| Decision Hierarchy           |   |           |
|------------------------------|---|-----------|
| Level 0                      | Level 1                                       | Glb Prio. |
| Datakwaliteit in procesmodel | Vertrouwd begrijpelijk niet complex 0.140     | 14.0%     |
|                              | Inzicht in bottlenecks 0.105                  | 10.5%     |
|                              | Bruikbaar als input systeemontwikkeling 0.064 | 6.4%      |
|                              | Zichtbaarheid dataflow 0.077                  | 7.7%      |
|                              | Transformatieproces duidelijk 0.106           | 10.6%     |
|                              | Zichtbaar waar DQ beïnvloed 0.096             | 9.6%      |
|                              | Zichtbaarheid kwaliteitscontroles 0.124       | 12.4%     |
|                              | Ondersteuning maatregelen DQ 0.066            | 6.6%      |
|                              | Meetbaarheid DQ product adhv dimensies 0.054  | 5.4%      |
|                              | Zichtbaarheid business impact 0.134           | 13.4%     |
|                              | Gedifferentieerde representatie data 0.034    | 3.4%      |
|                              |   | 1.0       |

## Bijlage 6 BPMN concepten gegevensgebruik

Tabel 25 Overzicht BPMN concepten gegevensgebruik

| Concept   | Naam                            | Toelichting  |
|---|---------------------------------|--|
|    | Task interactie met data store  | <p>Dit type task leest, creëert, bewerkt of verwijdt gegevens die zowel binnen als buiten de context van dit proces gebruikt worden.</p> <p>Aandachtspunt, data objecten zijn niet altijd expliciet gemodelleerd. Soms moet uit de context of beschrijving afgeleid worden dat gegevens gebruikt worden.</p> <p>Voorbeeld uit de steekproef, een van de stappen heet 'Registreren inkomend bericht in heffingsgebied'. Ook al zijn er geen data items te vinden in deze stap, uit de naam blijkt duidelijk dat er gegevens vastgelegd worden. Eén niveau dieper blijkt dat de gegevens ook gecontroleerd worden.</p> |
|   | Task interactie met data object | <p>Dit type task leest, creëert, bewerkt of verwijdt gegevens die alleen binnen de context van dit proces gebruikt worden.</p> <p>Hier geldt hetzelfde aandachtspunt als bij een task met interactie met een data store.</p>   |
|  | Send task                       | Dit type task zendt gegevens naar een ander proces of naar een participant buiten het proces.  |
|  | Receive task                    | Dit type task ontvangt gegevens van een ander proces of van een participant buiten het proces.   |
|  | Concepten met een message       | Een message kan voorkomen op een message flow, throwing en catching events (start, end, intermediate en boundary). Gegevens worden ontvangen van of verstuurd naar andere processen of participants buiten het proces.   |

## Bijlage 7 Feedback klankbordgroep

Onderstaande tabel geeft de bevindingen weer die naar voren zijn gekomen tijdens de bijeenkomsten met de klankbordgroep. Ook de resultaten van de proefbijeenkomst en de bespreking met de studiebegeleider zijn meegenomen, omdat deze waardevolle bijdragen opleverden voor het verbeteren van de methode. Van de één op één proefbijeenkomst is geen geluidsopname, daardoor zijn voor deze bevindingen geen tijdstippen opgenomen.

In de eerste kolom is weergegeven wat de bron is van de bevinding:

|     |   |
|-----|---|
| RH  | Eerste iteratie met afstudeerbegeleider Remko Helms van de OU.                                    |
| TG1 | Tweede iteratie, proefbijeenkomst met twee collega's die niet in de klankbordgroep zitten.        |
| TG2 | Tweede iteratie, proefbijeenkomst één op één met derde collega die niet in de klankbordgroep zit. |
| FG1 | Derde iteratie, focusgroepbijeenkomst van 3 december 2018 waarin drie collega's aanwezig waren.   |
| FG2 | Derde iteratie, focusgroepbijeenkomst van 6 december waarin de rest van de groep aanwezig was.    |

Idealiter zijn de resultaten van de iteraties eerst verwerkt voordat gestart wordt met de volgende iteratie. Dat is door tijdsdruk niet gelukt.

Tabel 26 Feedback klankbordgroep en verwerking

| Bron | Tijdstip | Nr.  | Bevinding/vraag/opmerking  | Verwerking   | Aanpassing methode  |
|------|----------|------|--|--|---|
| RH   | N.v.t.   | RH.1 | Bij toelichting overview opnemen dat er stappen zijn die het proces van bovenaf bekijken (architectuurprincipes) en dat scope daarbij het bedrijfsproces is, en stappen die van vanuit de inhoud van het werkproces kijken naar de individuele stappen. Toetsing op meerdere niveaus.          | Opmerking meegenomen in de presentatie van de methode voor de bijeenkomsten van de klankbordgroep. Ook in de toelichting opnemen.  | Ja, toelichting overview uitbreiden.                          |
| RH   | N.v.t.   | RH.2 | Stap 1, wat is de naam van concepten uit de eerste kolom?  | De bouwstenen van het metamodel van BPMN zijn concepten, dat is de naam die terug te vinden is op de site van de OMG, de organisatie die de taal beheert en ontwikkelt.  | Nee.  |
| RH   | N.v.t.   | RH.3 | Stap 1, hoe pak je dit systematisch aan en wat levert het op? Van links naar rechts? En met welk detailniveau? Is dan nog een definitie nodig van niveaus waarin het proces gedecomposeerd wordt? Is het resultaat een gemarkeerd proces? Is dit analoog aan de manier waarop je in stap 3 het | Systematiek toevoegen aan deze stap, maar zie ook onder meer TG.2 heroverweging van deze stap omdat bij organisaties waar deze methode toegepast kan worden bijna iedere stap wel met data te maken heeft. Doel stap zie TG.2. Naamgeving, mogelijk "verkenning". Wel toevoegen procesopologie uit Obers & Achterberg (2014) | Ja, andere uitwerking stap 1 en procesdecompositie toevoegen. |

| Bron       | Tijdstip | Nr.   | Bevinding/vraag/opmerking  | Verwerking   | Aanpassing methode   |
|------------|----------|-------|--|--|--|
|            |          |       | proces doorloopt?  | ketenproces, bedrijfsproces t/m activiteit.  |  |
| <b>RH</b>  | N.v.t.   | RH.4  | Stap 2, architectuurprincipes die genoemd zijn kun je (voor een aantal) alleen op het niveau van bedrijfsproces beschouwen. De andere stappen zoomen in op het werkproces.   | Meenemen in de beschrijving van stap 2. Verwijzen naar de definities van de decompositie en dat de andere lagen in de procesarchitectuur terug te vinden zijn.   | Ja, stap 2 uitbreiden.   |
| <b>RH</b>  | N.v.t.   | RH.5  | Stap 2, niet zozeer een principe, maar kun je aan de hand van visuele inspectie, dus vanuit de buitenkant van het proces, al bepaalde conclusies trekken? Zijn er bepaalde risicoprofielen op te stellen, bijvoorbeeld als datastromen bij elkaar komen of als de context van het gebruik verandert (koppelvlakken) en welke conclusies kun je daaruit afleiden? | Dat zou goed mogelijk kunnen zijn, maar is zonder verder onderzoek niet vast te stellen. Een mogelijke aanpak om dit te onderzoeken is de methode toe te passen voor meerdere processen en dan de processen die niet goed scoren te vergelijken op gemeenschappelijke kenmerken of plekken in het proces waar dataproblemen zich voordoen. | Nee, punt voor vervolgonderzoek.                                   |
| <b>RH</b>  | N.v.t.   | RH.6  | Onderscheid in stap 3 tussen R en B meer toelichten hoe je tot R of B komt. Ook onderscheid verder toelichten.   | Zie ook punt FG1.3, overwegen of het aangeven van een relatie niet volstaat.   | Ja, stap 3, overwegen of relatie volstaat.                         |
| <b>RH</b>  | N.v.t.   | RH.7  | Aanname bij zoals het nu beschreven is, is dat het proces loopt zoals het beschreven is. Requirements zijn een hulpmiddel, je zou ook kunnen kijken naar hoe het proces loopt in werkelijkheid.  | Ook klankbordgroep geeft aan dat meerdere bronnen nodig zijn, niet alleen model of requirements, denk aan interviews met beheerders en uitvoerders van het proces. Mogelijk kan dit meegenomen worden in de heroverweging van stap 1, bronnen inventariseren.  | Ja, stap 4, bronnen voor deze stap verbreden. Mogelijk ook stap 1. |
| <b>RH</b>  | N.v.t.   | RH.8  | Stap 7, extra aandachtspunt, toets of requirement niet toch al gerealiseerd is.  | Laatste stap in aantal substappen splitsen, daar kan deze in opgenomen worden.   | Ja, stap 7 splitsen in deelstappen.                                |
| <b>TG1</b> | 08:42    | TG1.1 | Doel is om problemen met datakwaliteit begrijpelijk te maken in termen van de business. Daar gebruik je BPMN voor. In hoeverre is die taal begrijpelijk voor de business?  | Is inderdaad een aandachtspunt. Medewerkers uit de business die de taal niet kennen zullen eerst de ruimte moeten krijgen om vertrouwd te raken met de taal of het proces moet zodanig gepresenteerd worden dat medewerkers meegenomen worden in de voor hen nieuwe taal. Meenemen in een nieuwe paragraaf van de methode, uitgangspunten. | Ja, uitgangspunten toevoegen.                                      |
| <b>TG1</b> | 19:34    | TG.2  | N.a.v. stap 1, in methode aanbeveling om op te sporen waar data wel gebruikt wordt maar niet gemodelleerd is. Opmerking Jos: soms kan het niet weergegeven worden omdat het teveel is. Je kan een item maar aan één UML klasse koppelen uit het CGM. Dag geeft beperkingen.  | Past in herbezinning op stap 1. Zie ook andere bevindingen RH.3. In alle processen van de onderzoeksorganisatie worden gegevens gebruikt. Doel van de stap heroverwegen. Manco's in kaart brengen, gegevensgebruik toetsen of oriënterend.   | Ja, stap 1, heroverweging doel en uitwerking stap 1.               |
| <b>TG1</b> | 24:51    | TG.3  | Bij stap 2 ook de voor dat project van toepassing zijnde architecturen meenemen en de principes die daar in genoemd staan, voor zover deze betrekking hebben op datakwaliteit. Architectuur kan de domeinarchitectuur zijn, maar ook referentiearchitectuur, concernarchitectuur etc.  | Meenemen in de tekst van stap 2.   | Ja, stap 2, opmerking toevoegen.                                   |
| <b>TG1</b> | 26:23    | TG.4  | Stap 2, zijn er naast de architectuurprincipes ook standaard requirements t.a.v. datakwaliteit die je mee kan nemen in de ontwikkeling van je proces? Of acceptatiecriteria van  | In de onderzoeksorganisatie niet kunnen vinden. Het past wel binnen het architectuurprincipe om processen efficiënt te ontwikkelen door hergebruik van standaard componenten. Dat is vooral de invulling van   | Nee.   |

| Bron       | Tijdstip           | Nr.   | Bevinding/vraag/opmerking  | Verwerking  | Aanpassing methode   |
|------------|--------------------|-------|--|---|--|
|            |                    |       | gegevens?  | het principe. Heeft geen impact op de methode.  |  |
| <b>TG1</b> | 20:53              | TG.5  | Draai de tabelweergave uit stap 3 3n 4 om, dus processtappen in de kolommen en kwaliteitsdimensies in de rijen, dan past hij 1 op 1 op de tabel onder het procesmodel in de grafische weergave (de notenbalk).   | Aanpassen.  | Ja, stap 3 en 4, tabelweergave aanpassen                               |
| <b>TG1</b> | 34:05              | TG.6  | Wat doe je als er geen procesmodel is of ontwerpdocumentatie in stap 4?  | Uitgangspunt is dat er een procesmodel is, opnemen in uitgangspunten. Zie ook andere bevindingen, naast ontwerpdocumentatie ook andere informatie nodig om deze stap uit te voeren.   | Ja, uitgangspunten toevoegen.  |
| <b>TG1</b> | 35:24              | TG.7  | Wat is relatie met het CGM in deze methode? Het is een procesgedreven aanpak, maar als je geen idee hebt wat de betekenis is van de gegevens die je behandelt in het proces, weet je dan voldoende? Met name als het gaat om koppelvlakken, andere context, moet je goede afspraken hebben en weten wat de gegevens betekenen. | Meenemen als uitgangspunt, de betekenis van de gegevens is beschreven in een CGM, en als onderdeel van de verkenning in stap 1.   | Ja, uitgangspunten toevoegen en opnemen bij stap 1.                    |
| <b>TG1</b> | 39:64<br>40:42     | TG.8  | Kwaliteitsdimensies kunnen verschillend geïnterpreteerd worden. Definitie opnemen wat met de dimensie bedoeld wordt. Soms zijn er ook verschillende definities in de literatuur te vinden, zoals bijvoorbeeld van tijdigheid. Maak expliciet welke je bedoelt.   | Meenemen in stap 3. Zie ook andere bevinding uit FG1 en FG2, betekenis kan zelfs verschillen per stap of per gevalscategorie.   | Ja, definities opnemen in stap 3.                                      |
| <b>TG1</b> | 46:48<br>48:51     | TG.9  | Inhoudelijk voor deze case, in dit proces worden gegevens ingewonnen voor verschillende doelen. Er zijn verschillende afnemers van de gegevens in dit proces. Kun je differentiëren naar afnemer bij het vastleggen van kwaliteitsstempel? Wat is je eenheid van beschouwing daarbij? De hele set of één exemplaar?            | Meenemen als requirement inhoudelijk in stap 2 naar aanleiding van principe quality at source.  | Nee, wel aanpassing inhoud uitwerking case bij stap 2.                 |
| <b>TG1</b> | 1:05:29            | TG.10 | Bottlenecks uit criteria, zijn dat bottlenecks in kwaliteit of in de doorstroming?   | Bedoeld is in de doorstroming. Criteria zijn al gewogen door leden klankbordgroep, maar zal tekst verduidelijken in rapport.  | Nee, wel aanvullende tekst bij criteria.                               |
| <b>TG1</b> | 1:18:24<br>1:19:05 | TG.11 | Voorwaarde voor deze aanpak is dat het proces overeenkomt met de werkelijkheid. Er zijn data mining tools waarmee het proces achteraf kan worden gereconstrueerd. Model mogelijk ijkten tegen de werkelijkheid.  | Zie ook TG.6, randvoorwaarde is niet alleen dat er een model beschikbaar is, maar ook dat het een afspiegeling is van de werkelijkheid. Meenemen in de uitgangspunten. Bij de uitgangspunten ook opnemen dat als er niet aan voldaan is, dat er eerst aan de uitgangspunten voldaan wordt. Dat zou dus in kunnen houden dat het model geactualiseerd moet worden. | Ja, aanvullen bij uitgangspunten en algemene tekst bij uitgangspunten. |
| <b>TG1</b> | 1:22:53            | TG.12 | Extra principe voor stap 2: enkelvoudig beheer, meervoudig gebruik.  | Toevoegen aan verzameling principes.  | Ja, principe toevoegen stap 2.   |
| <b>FG1</b> | 24:26              | FG.1  | Stap 2, bij het principe modelleren de hele keten. Wat is de hele keten als je zoals in dit proces verschillende afnemers bedient. Moet je alle processen van alle afnemers ook modelleren?  | Zie ook FG2 en RH.4 waarin dezelfde vraag gesteld wordt. Voor deze stap moet je de procesarchitectuur in ogenschouw nemen, niet alleen het werkproces waar je op inzoomt. In de procesarchitectuur zijn de  | Ja, stap 2 uitbreiden, bedrijfsproces                                  |

| Bron       | Tijdstip                      | Nr.   | Bevinding/vraag/opmerking   | Verwerking   | Aanpassing methode  |
|------------|-------------------------------|-------|---|--|---|
|            |                               |       |   | raakvlakken met afnemende processen weergegeven. Ook procestopologie toevoegen.  |   |
| <b>FG1</b> | 27:47                         | FG.2  | Bij stap 2, principe volledige keten, wanneer is je proces volledig weergegeven? Proces is volledig als alle data ingewonnen zijn.  | Klopt, is een inhoudelijke bevinding bij de case. Voor de methode geen effect. Wel enige samenhang met andere punten waarin aangegeven is dat het zinvol is om de dimensies te definiëren in stap 3.   | Nee.  |
| <b>FG1</b> | 29:55                         | FG.3  | In stap 3, is R en B nodig of volstaat een relatie? Wat is toegevoegde waarde van een R en een B?   | Zie ook RH.6.  | Ja, stap 3, alleen relatie.                               |
| <b>FG1</b> | 35:31                         | FG.4  | In stap 4 alleen de dimensies gebruiken die in stap 3 geselecteerd zijn.  | Toevoegen ter verduidelijking in stap 4.   | Ja, stap 4, toelichting uitbreiden                        |
| <b>FG1</b> | 38:56                         | FG.5  | Hoe ga je om met zaken die geconstateerd worden en die je accepteert? Bijv. een known error.  | Zie ook RH.8, meenemen in de detaillering van stap 7, prioritering en accepteren van bepaalde zaken die naar voren komen.  | Ja, stap 7 splitsen.                                      |
| <b>FG1</b> | 43:24                         | FG.6  | Wanneer krijgt het model nu zijn waarde?  | Waarde zit in requirements voor verbetering en de realisatie daarvan.  | Nee.  |
| <b>FG1</b> | 45:11                         | FG.7  | Aanpak heeft wel wat weg van de PRA-B analyse uit T-Map. Wordt op dit moment in de onderzoeksorganisatie niet meer uitgevoerd.  | Klopt.   | Nee.  |
| <b>FG1</b> | 52:08                         | FG.8  | Stap 3, goed om definities toe te voegen zodat iedereen hetzelfde beeld heeft van bijvoorbeeld volledigheid.  | Dit is ook in de testgroep en in de tweede bijeenkomst onderkend. Zie TG.8, definities toevoegen.  | Ja, stap 3, definities toevoegen.                         |
| <b>FG1</b> | 57:31                         | FG.9  | Wie scoort de dimensies in stap 4? Ervaring van Han is dat een subjectieve aangelegenheid is, soms beïnvloed door de politiek of door andere factoren. Verder wie betrek je er bij en hoe krijg je het zo objectief mogelijk? Vraagt een behoorlijke investering. | Toevoegen in stap 4 om score met relevante groep stakeholders uit te voeren. Als het lastig is om overeenstemming te bereiken met de groep ondanks beslisboom, dan suggestie om aanpak te volgen vergelijkbaar met pokeren in agile aanpak. Op die manier is de ervaring kom je relatief snel tot overeenstemming met een groep medewerkers. | Ja, stap 4 uitbreiden met stakeholders en aanpak.         |
| <b>FG1</b> | 59:10                         | FG.10 | Dit is geen eenmalige actie, inzicht verandert voortdurend en processen veranderen.   | Zie ook volgend punt FG.11, selectie van processen om deze methode op toe te passen.   | Ja, zie FG.11.  |
| <b>FG1</b> | 59:23                         | FG.11 | In vervolg op FG.7, analogie met PRA-B uit T-Map, je zou deze methode toe moeten passen op cruciale processen of processen waar risico groot is.  | Toevoegen bij inleiding op de methode, paragraaf over selectie van processen en criteria daarvoor. Vergelijk met criteria voor de case.  | Ja, inleiding uitbreiden                                  |
| <b>FG1</b> | 1:02:30<br>1:19:02            | FG.12 | Wees bewust van wat rode smileys doen met managers, die willen alles op groen hebben. Als je in het traject bepaalde risico's accepteert, laat die dan niet rood, maar geef die een andere kleur. Wit bijvoorbeeld.   | Meenemen in de aanpak, stap 7, fase van selecteren welke risico's je af wilt dekken en welke je accepteert. Geaccepteerde risico's wel zichtbaar maken maar met afwijkende kleur.  | Ja, stap 7 splitsen, en selectie toevoegen. Zie ook FG.5. |
| <b>FG1</b> | 1:06:22                       | FG.13 | Het procesmodel van deze case is een generiek model. Er lopen meerdere gevals categorieën doorheen. In hoeverre is de gevals categorie van invloed op de werking van het model?   | Bij bepalen van de scope kiezen waar model op toegepast gaat worden. Zie ook FG.11.  | Ja, inleiding uitbreiden over scope.                      |
| <b>FG1</b> | 1:06:22<br>1:28:30<br>1:30:32 | FG.14 | Meer bronnen nodig dan alleen procesmodel. Is procesmodel weergave van de werkelijkheid? Is het voor interviews nodig om stakeholders vast te stellen?  | Klopt, ook interviews etc. Toevoegen aan beslismodel stap 4. Ook meenemen: cijfers over de productie, bijvoorbeeld aantal incidenten en aantal verstoringen. Metingen uit Splunk. Mogelijk stakeholders in   | Ja, stap 4, toevoegen meerdere bronnen. Stakeholders      |

| Bron       | Tijdstip   | Nr.   | Bevinding/vraag/opmerking   | Verwerking  | Aanpassing methode   |
|------------|--|-------|---|---|--|
|            | 1:32:09<br>1:09:57                                 |       |   | kaart brengen.  | toevoegen.   |
| <b>FG1</b> | 1:08:16<br>1:15:55                                 | FG.15 | Volledigheid bij processtap opvangbak is een andere volledigheid dan bij de processtap logistiek. Hoe kan je dat laten zien. En kun je volledigheid geïsoleerd zien of moet je het afzetten ten opzichte van een cijfer van vorig jaar.   | Zie ook FG.8, definities toevoegen bij stap 3 en dat ook per processtap doen.   | Ja, stap 3, definities toevoegen, per processtap.                            |
| <b>FG1</b> | 1:15:01  | FG.16 | In de uitwerking van de case klopt er iets inhoudelijk niet. In het requirement staat dat het voor een melding geldt, maar het geldt voor totalen van de meldingen.   | Klopt, uitwerking aanpassen.  | Nee, niet methode, maar wel toepassing in case.                              |
| <b>FG1</b> | 1:23:03  | FG.17 | Stap 6 helpt om het effect van je maatregelen te bepalen.   | Meenemen in rationale van stap 6.   | Ja, stap 6, rationale uitbreiden.  |
| <b>FG1</b> | 1:23:47  | FG.18 | Op welke processen pas je deze methode toe? Wanneer is het relevant.  | Zie ook FG.11 en FG.13.   | Ja, inleiding uitbreiden over scope.   |
| <b>FG1</b> | 1:40:19  | FG.19 | In de tijd dat er waterval ontwikkeld werd bij de Belastingdienst werd Copafijth-BI als checklist gebruikt voor ontwerpers. Geeft dat nog aanknopingspunten?  | Na analyse, niet direct. Andersom zou datakwaliteit mogelijk als punt toegevoegd kunnen worden aan de checklist.  | Nee.   |
| <b>FG2</b> | 14:34  | FG.20 | Waarom deze case.   | Meenemen in inleiding, zie ook FG.11 en FG.13.  | Ja, inleiding.   |
| <b>FG2</b> | 18:52<br>21:31<br>22:43<br>52:20<br>54:08<br>54:42 | FG.21 | Discussie over definities van geloofwaardigheid en volledigheid en het belang van definities van de kwaliteitsdimensies. En ook concreet maken, hoe actueel is gevraagde actualiteit. Ook meenemen: wellicht heeft kwaliteitsdimensie betrekking op deel van de gegevensset in het proces. Dimensie kan per processtap andere lading krijgen. | Definities toevoegen, zie FG.15 en FG.8. Ook context meenemen van de dimensie (processtap en gebruikte gegevensset).  | Ja, stap 3, definities toevoegen, per processtap. Alleen afwijkende opnemen. |
| <b>FG2</b> | 25:25<br>26:45<br>35:09<br>55:56                   | FG.22 | Afnemers hebben ook allemaal eigen processen waarin de gegevens gebruikt worden. Wat breng je allemaal in kaart. De hele keten? En beschrijf je ook de eisen die de afnemers stellen?   | Zie FG.1, hele keten modelleren. Afnemende processen worden in BPMN vaak weergegeven als black box pool, dat wil zeggen dat je wel de deelnemer (participant) van het proces ziet, maar niet welke stappen bij die deelnemer uitgevoerd worden. Het model kan wel getoetst worden of de eisen die de afnemer stelt beschreven zijn. | Ja, stap 2, principe toevoegen dat eisen beschreven zijn van afnemers.       |
| <b>FG2</b> | 30:17  | FG.23 | De IPK voorschriften (interne kwaliteitsprocedure voor architectuurmodellen) kunnen helpen om de kwaliteit van modellen op orde te krijgen zodat niet gemodelleerd gegevensgebruik niet meer voorkomt.  | Klopt. Maar valt buiten scope van het onderzoek.  | Nee.   |
| <b>FG2</b> | 38:11<br>41:43                                     | FG.24 | Architectuurprincipe was altijd dat controle alleen bij de bron Inzicht veranderde, robuustheid boven efficiency, en ook partijen in de organisatie onafhankelijk van elkaar kunnen laten functioneren wordt als belangrijker gezien.   | Interessante discussie. Principe vooralsnog zo laten.   | Nee.   |
| <b>FG2</b> | 42:37  | FG.25 | Mag de afnemer de gegevens gebruiken? AVG meenemen in de principes?   | Binnen de Belastingdienst worden twee instrumenten gebruikt om dat te beoordelen: de WMK toets (Willen, Mogen, Kunnen) en de PIA  | Nee.   |

| Bron       | Tijdstip                      | Nr.   | Bevinding/vraag/opmerking  | Verwerking   | Aanpassing methode   |
|------------|-------------------------------|-------|--|--|--|
|            |                               |       |  | (Privacy Impact Analyse). Daarnaast worden er voorschriften ontwikkeld voor Privacy by Design. Buiten scope van dit onderzoek houden.  |  |
| <b>FG2</b> | 44:01                         | FG.26 | Hebben principes uit stap 2 te maken met gegevenskwaliteit?  | Ja, gebruikte bronnen zijn allemaal gerelateerd aan gegevenskwaliteit.   | Nee.   |
| <b>FG2</b> | 49:48                         | FG.27 | Is de R en de B in stap 3 van belang? Zou je ook kunnen volstaan met te zoeken naar lees, schrijf of bewerk?   | Zie ook RH.3 en TG.2.  | Ja, andere uitwerking stap 1 en procesdecompositie toevoegen.      |
| <b>FG2</b> | 50:50                         | FG.28 | Zou je dimensies nog willen wegen? Prioritering van dimensies?   | Zou toegevoegd kunnen worden om te bepalen welke requirements als eerste opgepakt moeten worden. Dat is een onderdeel in stap 7 (na aanpassing door de klankbordgroepen). Maar vermoedelijk spelen dan meer criteria een rol. Vooralsnog achterwege laten. | Nee.   |
| <b>FG2</b> | 53:57                         | FG.29 | Als je keuze maakt welke dimensies van belang zijn, dan wel alle dimensies zien, de volledige set.   | Toevoegen aan stap 3.  | Ja, stap 3, opmerking toevoegen dat volledige set meeneemt.        |
| <b>FG2</b> | 1:00:50                       | FG.30 | Wat zien we in het procesmodel, de Ist, de Soll? In architectuur zijn zowel baseline als target opgenomen in het model. Geldt dat voor procesmodel ook?  | Stap 4 laat de Ist situatie zien, stap 6 de Soll. Toevoegen aan de methode dat dit opgenomen moet worden in de documentatie en boven het model.  | Ja, stap 4 en stap 6, Ist en Soll toevoegen of Baseline en Target. |
| <b>FG2</b> | 1:07:18                       | FG.31 | Als ik een processtap bijvoorbeeld de volledigheid gecontroleerd wordt, beoordeel je in deze methode niet of die processtap zijn werk goed doet.   | Klopt gedeeltelijk. Naar aanleiding van andere opmerkingen wordt wel de gebruikte informatie uitgebreid om te oordelen in stap 4, bijvoorbeeld aantal verstoringen in productie. Dat zegt iets over hoe goed de stap wordt uitgevoerd.                     | Nee.   |
| <b>FG2</b> | 1:08:17                       | FG.32 | Beoordelen of het model volledig is, controle kan wel uitgevoerd worden, maar is er ook een foutpad opgenomen in het proces?   | Meenemen bij laatste architectuurprincipe waarin wordt gesteld dat er een proces moet zijn voor fouten in de output. Moet ook voor de throughput.  | Ja, stap 2, architectuurprincipe uitbreiden.                       |
| <b>FG2</b> | 1:11:01<br>1:11:57<br>1:14:32 | FG.33 | Maak van de beslisboom een DMN model met beslistabel. Past in huidige tooling en is een standaardtaal om beslissingen te modelleren. Kan daar ook bronnen in opnemen. Wellicht kun je stap 4 in zijn geheel in DMN ondersteunen zodat je een view kan genereren voor het resultaat van stap 4. | Goede suggestie, meenemen. View onderzoeken.   | Ja, stap 4, beslismodel ombouwen naar DMN.                         |
| <b>FG2</b> | 1:17:29<br>1:19:30            | FG.34 | Zou je in stap 1 ook aan kunnen geven welke gegevens verwerkt worden? Kanttekening: het proces in deze case is dan een lastig proces omdat er veel verschillende gegevens ingewonnen worden. Dat zou je op kunnen lossen met het modelleren van specialisaties in UML.                         | Meenemen in de heroriëntatie op stap 1, zie eerdere bevindingen zoals TG.2.  | Ja, stap 1, heroverweging doel en uitwerking stap 1.               |
| <b>FG2</b> | 1:22:39                       | FG.35 | In stap 2 verander je van blik, je beschouwt de hele keten en niet meer alleen de inhoud van het proces dat onderzocht   | Zie eerdere bevindingen zoals RH.4, in stap 2 hele keten meenemen, niveaus toelichten van procesarchitectuur.  | Ja, stap 2, procesdecompositie en                                  |



| Bron       | Tijdstip           | Nr.   | Bevinding/vraag/opmerking   | Verwerking  | Aanpassing methode  |
|------------|--------------------|-------|---|---|---|
|            |                    |       | wordt.  |   | blikveld toevoegen  |
| <b>FG2</b> | 1:27:27            | FG.36 | Voor het archiveren (bevinding uit de case) zijn referentiearchitecturen beschikbaar.   | Goed punt, maar wel specifiek voor de Belastingdienst.  | Nee.  |
| <b>FG2</b> | 1:29:40<br>1:33:07 | FG.37 | In stap 2 een principe opnemen met betrekking tot goed ontwerpen? De principes zijn er wel, maar ze worden niet altijd toegepast zoals ook blijkt uit de case.                  | Er zijn ontwikkelingen binnen de Belastingdienst dat er strakker gestuurd gaat worden op proceskwaliteit, of beter, de kwaliteit van het model. Buiten de scope van de methode houden, kwaliteit van een model is een heel breed begrip. Daar waar het gegevenskwaliteit raakt meenemen als aandachtspunt zoals dat nu gedaan is. | Nee.  |
| <b>FG2</b> | 1:34:47            | FG.38 | Niet alleen dimensie definiëren, maar ook waar het betrekking op heeft.   | Zie eerdere bevindingen, meenemen in stap 3.  | Ja, stap 3, definities toevoegen en gevalscategorie.                                  |
| <b>FG2</b> | 1:38:44            | FG.39 | Hoe laat je in stap 4 zien dat je meerdere bronnen hebt gebruikt? Doe je dat door de bronnen te gebruiken om het model te valideren? Of verwerk je de bronnen in het resultaat? | Meenemen in de uitwerking van stap 4  | Ja, stap 4, meerdere bronnen meenemen in resultaat.                                   |
| <b>FG2</b> | 1:40:30            | FG.40 | Hoe ga je om met geaccepteerde risico's in de Soll.   | Zie FG.12   | Ja, stap 7 splitsen, en selectie toevoegen. Zie ook FG.5.                             |
| <b>FG2</b> | 1:45:52            | FG.41 | Stap 6, laat dit zien welke smileys weer groen gaan worden?   | Ja, meenemen in de beschrijving, relatie tussen Ist en Soll. Zie ook FG.30, uitbreiden met relatie voor de notenbalk.   | Ja, stap 4 en stap 6, Ist en Soll toevoegen of Baseline en Target en gevolgen scores. |
| <b>FG2</b> | 1:47:11            | FG.42 | Stap 6 kan ook tot gevolg hebben dat er een processtap toegevoegd moet worden. Kan dat?   | Ja, requirements kunnen zeker gevolgen hebben voor het proces. Als opmerking toevoegen bij stap 5.  | Ja, stap 5, opmerking toevoegen ter verduidelijking.                                  |
| <b>FG2</b> | 1:47:38            | FG.43 | Requirements integreren in transitie architectuur en werkpakketten? Voordeel dat je het dan uit kan zetten in de tijd.  | Goede suggestie, maar buiten scope methode houden.  | Nee.  |
| <b>FG2</b> | 1:49:13            | FG.44 | Kun je deze methode koppelen aan een risicomodel? Zodat de business kan wegen welk risico gelopen wordt als een dimensie gemist wordt.  | Ja, meenemen in stap 7, zie ook FG.40, FG.12 en FG.5  | Ja, stap 7, risicomodel toevoegen bij afweging.                                       |

## Bijlage 8 Vragenlijst klankbordgroep

Onderstaande tabel geeft de criteria weer zoals die gewogen zijn door de deelnemers. Per criterium zijn vragen opgesteld om te meten hoe de deelnemers het model beoordelen op dit criterium.

Tabel 27 Overzicht met vragen per criterium t.b.v. vragenlijst sessie focusgroep

| Criterium   | Vragen  |
|---|---|
| Modellen moeten vertrouwd zijn voor stakeholders, begrijpelijk zijn en niet te complex.                                   | <ol style="list-style-type: none"> <li>1. De stappen in de methode zijn mij duidelijk.</li> <li>2. De overview van de methode laat mij zien wat de methode mij gaat opleveren. <i>(afbeelding toevoegen)</i></li> <li>3. De grafische weergave van stap 4 laat mij zien welke stappen in het proces aandacht behoeven als het om datakwaliteit gaat. <i>(afbeelding toevoegen)</i></li> <li>4. De tabel uit stap 4 waarin de processtappen afgezet zijn tegen de kwaliteitsdimensies geeft mij inzicht in de sterke en zwakke plekken van het proces als het om datakwaliteit gaat. <i>(afbeelding toevoegen)</i></li> <li>5. Het model waarin de requirements voor datakwaliteit zijn weergegeven geeft mij inzicht in de opgestelde requirements. <i>(afbeelding toevoegen)</i></li> <li>6. De tabel waarin requirements zijn afgezet tegen processtappen geeft mij meer inzicht in waar in het proces welke requirements gesteld worden. <i>(afbeelding toevoegen)</i></li> <li>7. Het nut van de stappen in de methode is mij duidelijk.</li> <li>8. Het is mij duidelijk welke stappen in de methode gevolgd moeten worden.</li> </ol> |
| Het model moet inzicht geven in bottlenecks in het proces.  | <ol style="list-style-type: none"> <li>9. Het model dat na toepassing van de methode ontstaat, laat mij zien waar de bottlenecks in het proces zitten.</li> </ol>   |
| De opgeleverde procesmodellen moeten bruikbaar zijn als input voor het systeemontwikkelingstraject.                       | <ol style="list-style-type: none"> <li>10. Het resultaat van de methode kan ik als input gebruik voor het IV voortbrengingstraject.</li> <li>11. Ik kan het resultaat van stap 2 (toetsen model op architectuurprincipes) en stap 5 (vaststellen requirements datakwaliteit) gebruiken om features en story's op te stellen voor de backlog. <i>(afbeelding toevoegen)</i></li> <li>12. De requirements uit stap 5 sluiten goed aan bij de features en story's die onze IV voortbrenging kent.</li> </ol>   |
| Het model moet de raw dataflow, component dataflow en IP-flow kunnen laten zien.  | <ol style="list-style-type: none"> <li>13. Het procesmodel uit stap 4 van de methode laat zien welke brongegevens gebruikt worden die leiden tot het informatie-eindproduct. <i>(afbeelding toevoegen)</i></li> </ol>   |
| Het transformatieproces moet duidelijk zijn.  | <ol style="list-style-type: none"> <li>14. Het procesmodel uit stap 4 van de methode laat zien hoe brongegevens getransformeerd worden tot het eindproduct.</li> </ol>  |
| De belangrijkste fasen in het proces en kritieke fasen die datakwaliteit beïnvloeden moeten gevisualiseerd kunnen worden. | <ol style="list-style-type: none"> <li>15. Toepassing van de methode laat mij zien waar de kritieke fasen in het proces zitten die datakwaliteit beïnvloedt.</li> </ol>   |
| Kwaliteitscontroles zijn expliciet zichtbaar.   | <ol style="list-style-type: none"> <li>16. De methode laat zien waar in het proces kwaliteitscontroles opgenomen zijn.</li> </ol>   |
| Het model moet ondersteunend zijn in het implementeren van kwaliteitsmaatregelen.   | <ol style="list-style-type: none"> <li>17. Het resultaat van de methode helpt mij in het implementeren van kwaliteitsmaatregelen zoals bijvoorbeeld het uitvoeren van een extra controle.</li> </ol>  |
| Het model moet het mogelijk maken om op verschillende abstractieniveaus de  | <ol style="list-style-type: none"> <li>18. Het resultaat van de methode laat mij zien wat de kwaliteit van het eindproduct is aan de hand van de relevante kwaliteitsdimensies.</li> </ol>  |

| Criterium   | Vragen  |
|---|---|
| kwaliteit van het informatieproduct meetbaar te maken aan de hand van de relevante kwaliteitsdimensies.                           |   |
| De methode moet concepten bevatten om het belang van de business impact en waarde van datakwaliteit te kunnen laten zien.         | 19. De methode laat zien wat het belang van de impact op de business is van de kwaliteit van de data. |
| De methode moet rekening houden met de relevante context van data zodat een gedifferentieerde representatie van data mogelijk is. | 20. De methode houdt rekening met de relevante context van data die gebruikt worden in het model.     |

De vragen zijn overgenomen in de tool die beschikbaar gesteld is op [www.enquetesmaken.com](http://www.enquetesmaken.com).

De gegenereerde lijst staat hierna. Om de vragen te scoren is een Likert schaal gebruikt. Voordat de bijeenkomst met de klankbordgroep plaatsvond, is de opzet van de bijeenkomst en de vragenlijst in kleine setting getest met een drietal collega's. Deze bijeenkomst vond plaats op maandag 26 november met twee leden van de testgroep. Met de derde heeft op 27 november een één op één afspraak plaats gevonden. Tijdens de bijeenkomst zijn tips gegeven voor de presentatie, het toevoegen van een afbeelding, een andere volgorde, dat soort zaken. Ook bleek dat er te weinig tijd was om de vragenlijst tijdens de bijeenkomst in te vullen. Dit hebben de deelnemers na afloop gedaan. Daarnaast leverde deze sessie de nodige inhoudelijke bevindingen die meegenomen zijn in de tweede iteratie van de methode.

## Introductie

Dit onderzoek is onderdeel van mijn afstudeeropdracht aan de Open Universiteit en gaat over datakwaliteit in procesmodellen. Aan de hand van criteria uit de literatuur wil ik de ontworpen methode toetsen. Als er vragen niet duidelijk zijn of verder toelichting behoeven laat het dan even weten. In de handout staan de figuren opgenomen die horen bij een aantal van de vragen.

## Pagina 2

**De stappen in de methode zijn mij duidelijk. \***

☐ Helemaal mee eens
 ☐ Mee eens
 ☐ Neutraal
 ☐ Mee oneens
 ☐ Helemaal mee oneens

**Het model dat na toepassing van de methode ontstaat, laat mij zien waar de bottlenecks in het proces zitten. \***

☐ Helemaal mee eens
 ☐ Mee eens
 ☐ Neutraal
 ☐ Mee oneens
 ☐ Helemaal mee oneens

**De overview van de methode laat mij zien wat de methode mij gaat opleveren. Zie afbeelding 1 handout. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**Het resultaat van de methode kan ik als input gebruik voor het IV voortbrengingstraject. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**De grafische weergave van stap 4 laat mij zien welke stappen in het proces aandacht behoeven als het om datakwaliteit gaat. Zie afbeelding 2 handout. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**Ik kan het resultaat van stap 2 (toetsen model op architectuurprincipes) en stap 5 (vaststellen requirements datakwaliteit) gebruiken om features en story's op te stellen voor de backlog. Zie afbeelding 3 handout. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**De tabel uit stap 4 waarin de processtappen afgezet zijn tegen de kwaliteitsdimensies geeft mij inzicht in de sterke en zwakke plekken van het proces als het om datakwaliteit gaat. Zie afbeelding 4 handout. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**Het model waarin de requirements voor datakwaliteit zijn weergegeven geeft mij inzicht in de opgestelde requirements. Zie afbeelding 5 handout. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**De requirements uit stap 5 sluiten goed aan bij de features en story's die onze IV voortbrenging kent. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**De tabel waarin requirements zijn afgezet tegen processtappen geeft mij meer inzicht in waar in het proces welke requirements gesteld worden. Zie afbeelding 6 handout. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**Het procesmodel uit stap 4 van de methode laat zien welke brongegevens gebruikt worden die leiden tot het informatie-eindproduct. Zie afbeelding 2 handout. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**Het nut van de stappen in de methode is mij duidelijk. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**Toepassing van de methode laat mij zien waar de kritieke fasen in het proces zitten die datakwaliteit beïnvloedt. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**De methode laat zien waar in het proces kwaliteitscontroles opgenomen zijn. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**Het procesmodel uit stap 4 van de methode laat zien hoe brongegevens getransformeerd worden tot het eindproduct. Zie afbeelding 2 handout. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**Het is mij duidelijk welke stappen in de methode gevolgd moeten worden. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**Het resultaat van de methode helpt mij in het implementeren van kwaliteitsmaatregelen zoals bijvoorbeeld het uitvoeren van een extra controle. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**Het resultaat van de methode laat mij zien wat de kwaliteit van het eindproduct is aan de hand van de relevante kwaliteitsdimensies. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**De methode laat zien wat het belang van de impact op de business is van de kwaliteit van de data. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

**De methode houdt rekening met de relevante context van data die gebruikt worden in het model. \***

☐ Helemaal mee eens   ☐ Mee eens   ☐ Neutraal   ☐ Mee oneens   ☐ Helemaal mee oneens

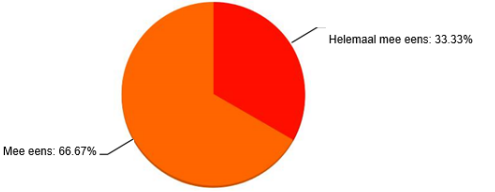
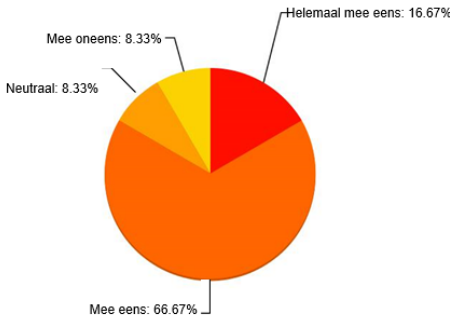
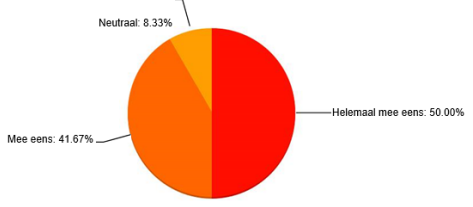
## **Afsluiting**

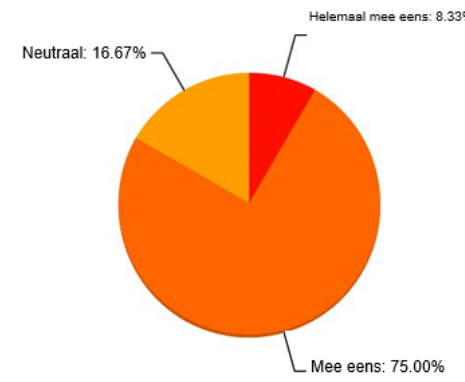
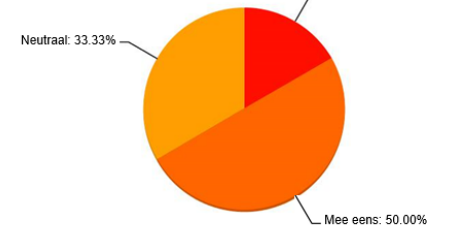
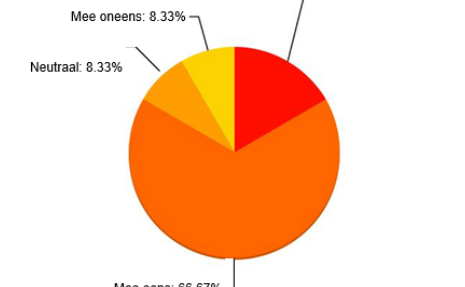
Hartelijk bedankt voor de medewerking! Hierna volgt nog een pagina van [www.enquetesmaken.com](http://www.enquetesmaken.com) waar deze vragenlijst in gemaakt is. Deze pagina kun je afsluiten.

» **Redirection to final page of Enquêtes Maken** ([wijzigen](#))

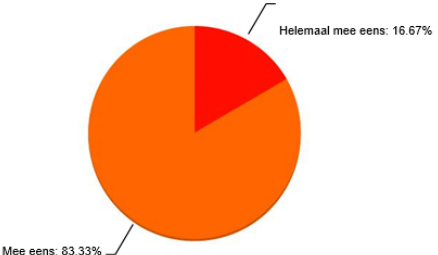
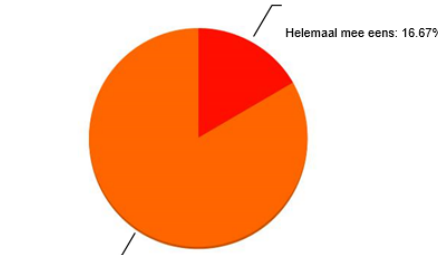
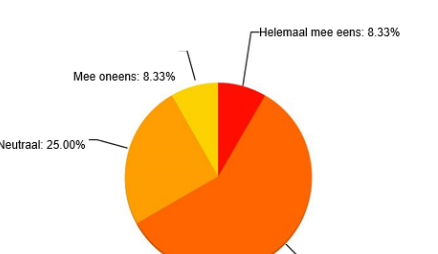
## Bijlage 9 Antwoorden vragenlijst klankbordgroep

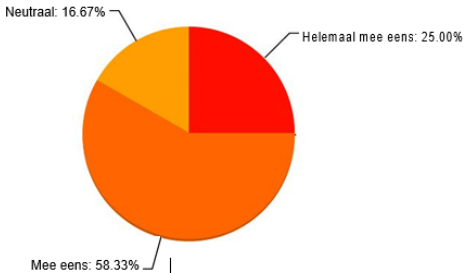
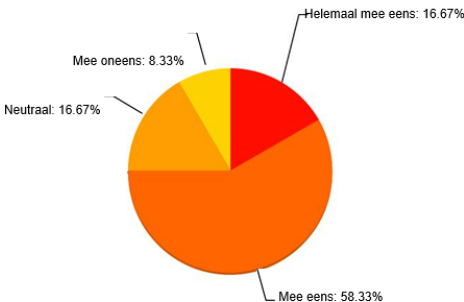
Tabel 28 Antwoorden vragenlijst en verwerking

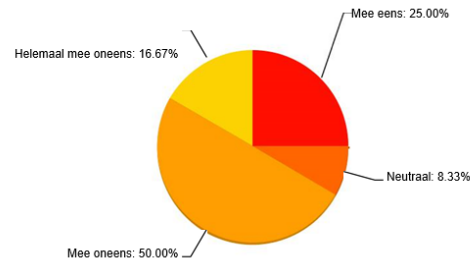
| Criterium en weging   | Vragen   | Resultaat   | Verwerking   |
|---|--|---|--|
| <b>Modellen moeten vertrouwd zijn voor stakeholders, begrijpelijk zijn en niet te complex. Weging 14%</b> | 1. De stappen in de methode zijn mij duidelijk.  |  <p>Mee eens: 66.67%<br/>Helemaal mee eens: 33.33%</p>  | Iedereen vindt dat de stappen duidelijk zijn. Geen aanpassing nodig.                                 |
|   | 2. De overview van de methode laat mij zien wat de methode mij gaat opleveren. <i>(afbeelding toevoegen)</i>   |  <p>Mee oneens: 8.33%<br/>Neutraal: 8.33%<br/>Helemaal mee eens: 16.67%<br/>Mee eens: 66.67%</p> | De meerderheid vindt dat de overview laat zien wat de methode gaat opleveren. Geen aanpassing nodig. |
|   | 3. De grafische weergave van stap 4 laat mij zien welke stappen in het proces aandacht behoeven als het om datakwaliteit gaat. <i>(afbeelding toevoegen)</i> |  <p>Neutraal: 8.33%<br/>Mee eens: 41.67%<br/>Helemaal mee eens: 50.00%</p>                      | Iedereen is het eens met deze stelling met één neutrale reactie. Geen aanpassing nodig.              |

| Criterium en weging | Vragen   | Resultaat  | Verwerking |            |          |        |                   |        |                   |        |  |       |  |
|---------------------|--|--|------------|------------|----------|--------|-------------------|--------|-------------------|--------|--|-------|--|
|                     | 4. De tabel uit stap 4 waarin de processtappen afgezet zijn tegen de kwaliteitsdimensies geeft mij inzicht in de sterke en zwakke plekken van het proces als het om datakwaliteit gaat. (afbeelding toevoegen) |  <table border="1"><thead><tr><th>Reactie</th><th>Percentage</th></tr></thead><tbody><tr><td>Mee eens</td><td>75.00%</td></tr><tr><td>Neutraal</td><td>16.67%</td></tr><tr><td>Helemaal mee eens</td><td>8.33%</td></tr></tbody></table>   | Reactie    | Percentage | Mee eens | 75.00% | Neutraal          | 16.67% | Helemaal mee eens | 8.33%  | De meerderheid is het eens met deze stelling. Geen aanpassing nodig. |       |  |
| Reactie             | Percentage   |  |            |            |          |        |                   |        |                   |        |  |       |  |
| Mee eens            | 75.00%   |  |            |            |          |        |                   |        |                   |        |  |       |  |
| Neutraal            | 16.67%   |  |            |            |          |        |                   |        |                   |        |  |       |  |
| Helemaal mee eens   | 8.33%  |  |            |            |          |        |                   |        |                   |        |  |       |  |
|                     | 5. Het model waarin de requirements voor datakwaliteit zijn weergegeven geeft mij inzicht in de opgestelde requirements. (afbeelding toevoegen)  |  <table border="1"><thead><tr><th>Reactie</th><th>Percentage</th></tr></thead><tbody><tr><td>Mee eens</td><td>50.00%</td></tr><tr><td>Neutraal</td><td>33.33%</td></tr><tr><td>Helemaal mee eens</td><td>16.67%</td></tr></tbody></table>  | Reactie    | Percentage | Mee eens | 50.00% | Neutraal          | 33.33% | Helemaal mee eens | 16.67% | De meerderheid is het eens met deze stelling. Geen aanpassing nodig. |       |  |
| Reactie             | Percentage   |  |            |            |          |        |                   |        |                   |        |  |       |  |
| Mee eens            | 50.00%   |  |            |            |          |        |                   |        |                   |        |  |       |  |
| Neutraal            | 33.33%   |  |            |            |          |        |                   |        |                   |        |  |       |  |
| Helemaal mee eens   | 16.67%   |  |            |            |          |        |                   |        |                   |        |  |       |  |
|                     | 6. De tabel waarin requirements zijn afgezet tegen processtappen geeft mij meer inzicht in waar in het proces welke requirements gesteld worden. (afbeelding toevoegen)  |  <table border="1"><thead><tr><th>Reactie</th><th>Percentage</th></tr></thead><tbody><tr><td>Mee eens</td><td>66.67%</td></tr><tr><td>Helemaal mee eens</td><td>16.67%</td></tr><tr><td>Mee oneens</td><td>8.33%</td></tr><tr><td>Neutraal</td><td>8.33%</td></tr></tbody></table> | Reactie    | Percentage | Mee eens | 66.67% | Helemaal mee eens | 16.67% | Mee oneens        | 8.33%  | Neutraal   | 8.33% | De meerderheid is het eens met deze stelling. In die zin geen aanpassing nodig. In de toelichting geeft een van de respondenten aan om de manier waarop requirements opgesteld worden aan te laten sluiten bij de manier die gebruikelijk is in de organisatie en het project. Suggestie meenemen bij de uitgangspunten. |
| Reactie             | Percentage   |  |            |            |          |        |                   |        |                   |        |  |       |  |
| Mee eens            | 66.67%   |  |            |            |          |        |                   |        |                   |        |  |       |  |
| Helemaal mee eens   | 16.67%   |  |            |            |          |        |                   |        |                   |        |  |       |  |
| Mee oneens          | 8.33%  |  |            |            |          |        |                   |        |                   |        |  |       |  |
| Neutraal            | 8.33%  |  |            |            |          |        |                   |        |                   |        |  |       |  |

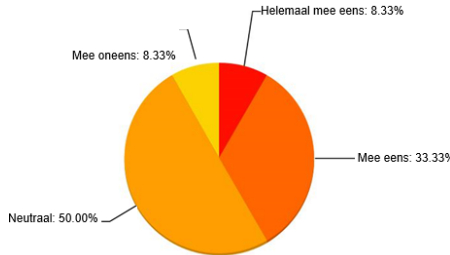
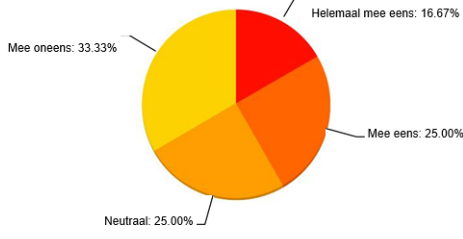


| Criterium en weging  | Vragen  | Resultaat  | Verwerking  |
|--|---|--|---|
|  | 7. Het nut van de stappen in de methode is mij duidelijk.   |  <p>Mee eens: 83.33%</p> <p>Helemaal mee eens: 16.67%</p>  | Iedereen is het met deze stelling eens. Geen aanpassing nodig.  |
|  | 8. Het is mij duidelijk welke stappen in de methode gevolgd moeten worden.                                      |  <p>Mee eens: 83.33%</p> <p>Helemaal mee eens: 16.67%</p>  | <p>Iedereen is het met deze stelling eens. Geen aanpassing nodig.</p> <p><b>Conclusie criterium vertrouwd, begrijpelijk, niet te complex:</b> de methode voldoet aan dit criterium.</p>   |
| Het model moet inzicht geven in bottlenecks in het proces.<br>Weging 10,5% | 9. Het model dat na toepassing van de methode ontstaat, laat mij zien waar de bottlenecks in het proces zitten. |  <p>Mee eens: 58.33%</p> <p>Neutraal: 25.00%</p> <p>Mee oneens: 8.33%</p> <p>Helemaal mee eens: 8.33%</p> | <p>De meerderheid is het hier mee eens, 66,66%. Echter, aandeel neutraal en oneens is een derde. In de open reacties staat dat het model niet de bottlenecks in doorlooptijd en verwerking laat zien.</p> <p><b>Analyse:</b> Een procesmodel laat de stappen zien die nodig zijn om een informatieproduct te maken. Bij de onderzoeksorganisatie is het verplicht om aantallen en doorlooptijd per stap op te geven. Dat helpt om een beeld te krijgen van bottlenecks. Echter, voor een beter beeld is het ook nodig om naar de uitvoering van het proces te kijken. Roser, Lorentzen &amp; Deuse (2014) bijvoorbeeld nemen het observeren van processen en voorraadpunten mee in hun analyse van bottlenecks.</p> |

| Criterium en weging  | Vragen   | Resultaat   | Verwerking   |                   |                   |          |          |          |          |            |  |   |
|--|--|---|--|-------------------|-------------------|----------|----------|----------|----------|------------|--|---|
|  |  |   | <b>Conclusie:</b> Voor deze methode is het zichtbaar maken van bottlenecks voldoende geborgd. Ligt het accent van een procesanalyse op het vinden van bottlenecks, dan zijn aanvullende stappen nodig zoals het observeren van procesuitvoering en voorraadpunten. |                   |                   |          |          |          |          |            |  |   |
| De opgeleverde procesmodellen moeten bruikbaar zijn als input voor het systeemontwikkelingstraject.<br>Weging 6,4%   | 10. Het resultaat van de methode kan ik als input gebruik voor het IV voortbrengingstraject.   |  <table border="1"><thead><tr><th>Reactie</th><th>Percentage</th></tr></thead><tbody><tr><td>Helemaal mee eens</td><td>25.00%</td></tr><tr><td>Mee eens</td><td>58.33%</td></tr><tr><td>Neutraal</td><td>16.67%</td></tr></tbody></table> | Reactie  | Percentage        | Helemaal mee eens | 25.00%   | Mee eens | 58.33%   | Neutraal | 16.67%     | De meerderheid is het eens met deze stelling. Geen aanpassing nodig. |   |
|  | Reactie  | Percentage  |  |                   |                   |          |          |          |          |            |  |   |
| Helemaal mee eens  | 25.00%   |   |  |                   |                   |          |          |          |          |            |  |   |
| Mee eens   | 58.33%   |   |  |                   |                   |          |          |          |          |            |  |   |
| Neutraal   | 16.67%   |   |  |                   |                   |          |          |          |          |            |  |   |
| 11. Ik kan het resultaat van stap 2 (toetsen model op architectuurprincipes) en stap 5 (vaststellen requirements datakwaliteit) gebruiken om features en story's op te stellen voor de backlog. (afbeelding toevoegen) |  <table border="1"><thead><tr><th>Reactie</th><th>Percentage</th></tr></thead><tbody><tr><td>Helemaal mee eens</td><td>16.67%</td></tr><tr><td>Mee eens</td><td>58.33%</td></tr><tr><td>Neutraal</td><td>16.67%</td></tr><tr><td>Mee oneens</td><td>8.33%</td></tr></tbody></table> | Reactie   | Percentage   | Helemaal mee eens | 16.67%            | Mee eens | 58.33%   | Neutraal | 16.67%   | Mee oneens | 8.33%  | De meerderheid is het eens met deze stelling. Echter, aandeel neutraal en oneens is een derde.<br><b>Analyse:</b> In de open reacties wordt verder geen toelichting gegeven. In de klankbordbijeenkomsten is wel gesteld dat de requirements an sich niet voldoende zijn. In stap 7 is nog prioritering nodig. Dat wordt toegevoegd aan de methode.<br><b>Conclusie:</b> stap 7 uitbreiden zoals in feedback klankbordgroep meegegeven. |
| Reactie  | Percentage   |   |  |                   |                   |          |          |          |          |            |  |   |
| Helemaal mee eens  | 16.67%   |   |  |                   |                   |          |          |          |          |            |  |   |
| Mee eens   | 58.33%   |   |  |                   |                   |          |          |          |          |            |  |   |
| Neutraal   | 16.67%   |   |  |                   |                   |          |          |          |          |            |  |   |
| Mee oneens   | 8.33%  |   |  |                   |                   |          |          |          |          |            |  |   |

| Criterium en weging   | Vragen   | Resultaat  | Verwerking |            |            |        |          |        |                     |        |                   |       |   |
|---|--|--|------------|------------|------------|--------|----------|--------|---------------------|--------|-------------------|-------|---|
|   | 12. De requirements uit stap 5 sluiten goed aan bij de features en story's die onze IV voortbrenging kent.   |  <table><thead><tr><th>Reactie</th><th>Percentage</th></tr></thead><tbody><tr><td>Neutraal</td><td>41.67%</td></tr><tr><td>Mee eens</td><td>33.33%</td></tr><tr><td>Mee oneens</td><td>16.67%</td></tr><tr><td>Helemaal mee eens</td><td>8.33%</td></tr></tbody></table>     | Reactie    | Percentage | Neutraal   | 41.67% | Mee eens | 33.33% | Mee oneens          | 16.67% | Helemaal mee eens | 8.33% | <p>Een groot deel geeft neutraal aan bij deze vraag, een minderheid is het eens is met deze stelling.</p> <p><b>Analyse:</b> In de open reacties wordt verder geen toelichting gegeven. Ook in de klankbordsessies is hier verder geen opmerking over gemaakt, alleen dat de manier waarop requirements uitgewerkt worden het beste aan kan sluiten op de manier die in de organisatie gebruikelijk is.</p> <p><b>Conclusie:</b> opnemen als punt voor verder onderzoek.</p> <p><b>Conclusie criterium bruikbaarheid systeemontwikkelingstraject:</b> over het algemeen voldoende, kleine aanpassingen in stap 7 en fine tuning in vervolgonderzoek</p>                                 |
| Reactie   | Percentage   |  |            |            |            |        |          |        |                     |        |                   |       |   |
| Neutraal  | 41.67%   |  |            |            |            |        |          |        |                     |        |                   |       |   |
| Mee eens  | 33.33%   |  |            |            |            |        |          |        |                     |        |                   |       |   |
| Mee oneens  | 16.67%   |  |            |            |            |        |          |        |                     |        |                   |       |   |
| Helemaal mee eens   | 8.33%  |  |            |            |            |        |          |        |                     |        |                   |       |   |
| Het model moet de raw dataflow, component dataflow en IP-flow kunnen laten zien.<br>Weging 7,7% | 13. Het procesmodel uit stap 4 van de methode laat zien welke brongegevens gebruikt worden die leiden tot het informatie-eindproduct. (afbeelding toevoegen) |  <table><thead><tr><th>Reactie</th><th>Percentage</th></tr></thead><tbody><tr><td>Mee oneens</td><td>50.00%</td></tr><tr><td>Mee eens</td><td>25.00%</td></tr><tr><td>Helemaal mee oneens</td><td>16.67%</td></tr><tr><td>Neutraal</td><td>8.33%</td></tr></tbody></table>   | Reactie    | Percentage | Mee oneens | 50.00% | Mee eens | 25.00% | Helemaal mee oneens | 16.67% | Neutraal          | 8.33% | <p>De meerderheid is het eens met deze stelling.</p> <p><b>Analyse:</b> In de open vragen is geen toelichting te vinden op de twee reacties die het niet eens waren met de stelling.</p> <p><b>Conclusie:</b> Geen aanpassing nodig.</p>  |
| Reactie   | Percentage   |  |            |            |            |        |          |        |                     |        |                   |       |   |
| Mee oneens  | 50.00%   |  |            |            |            |        |          |        |                     |        |                   |       |   |
| Mee eens  | 25.00%   |  |            |            |            |        |          |        |                     |        |                   |       |   |
| Helemaal mee oneens   | 16.67%   |  |            |            |            |        |          |        |                     |        |                   |       |   |
| Neutraal  | 8.33%  |  |            |            |            |        |          |        |                     |        |                   |       |   |
| Het transformatieproces moet duidelijk zijn.<br>Weging 10,6%                                    | 14. Het procesmodel uit stap 4 van de methode laat zien hoe brongegevens getransformeerd worden tot het eindproduct.   |  <table><thead><tr><th>Reactie</th><th>Percentage</th></tr></thead><tbody><tr><td>Mee oneens</td><td>50.00%</td></tr><tr><td>Mee eens</td><td>25.00%</td></tr><tr><td>Helemaal mee oneens</td><td>16.67%</td></tr><tr><td>Neutraal</td><td>8.33%</td></tr></tbody></table> | Reactie    | Percentage | Mee oneens | 50.00% | Mee eens | 25.00% | Helemaal mee oneens | 16.67% | Neutraal          | 8.33% | <p>De meerderheid is niet eens met deze stelling.</p> <p><b>Analyse:</b> Vermoedelijk komt dat doordat het procesmodel in lagen op gebouwd is. Het model dat in het onderzoek gebruikt is geeft het proces weer op processtapniveau. Op dat niveau is de gegevensbewerking niet zichtbaar. Dat is pas zichtbaar op het niveau van activiteit. Dit niveau is met de deelnemers van de klankbordgroep niet besproken.</p> <p><b>Conclusie:</b> deze bevinding behoeft verder onderzoek. Als er tijd is in een volgende iteratie meenemen, anders opnemen als punt voor vervolgonderzoek. Het gewicht van het criterium geeft aan dat de klankbordgroep dit een belangrijk punt vindt.</p> |
| Reactie   | Percentage   |  |            |            |            |        |          |        |                     |        |                   |       |   |
| Mee oneens  | 50.00%   |  |            |            |            |        |          |        |                     |        |                   |       |   |
| Mee eens  | 25.00%   |  |            |            |            |        |          |        |                     |        |                   |       |   |
| Helemaal mee oneens   | 16.67%   |  |            |            |            |        |          |        |                     |        |                   |       |   |
| Neutraal  | 8.33%  |  |            |            |            |        |          |        |                     |        |                   |       |   |

| Criterium en weging  | Vragen   | Resultaat  | Verwerking |            |          |        |                   |        |          |        |            |       |   |
|--|--|--|------------|------------|----------|--------|-------------------|--------|----------|--------|------------|-------|---|
| <b>De belangrijkste fasen in het proces en kritieke fasen die datakwaliteit beïnvloeden moeten gevisualiseerd kunnen worden.</b><br><b>Weging 9,6%</b> | 15. Toepassing van de methode laat mij zien waar de kritieke fasen in het proces zitten die datakwaliteit beïnvloedt.                              | <table border="1"><thead><tr><th>Reactie</th><th>Percentage</th></tr></thead><tbody><tr><td>Mee eens</td><td>58.33%</td></tr><tr><td>Helemaal mee eens</td><td>16.67%</td></tr><tr><td>Neutraal</td><td>16.67%</td></tr><tr><td>Mee oneens</td><td>8.33%</td></tr></tbody></table> | Reactie    | Percentage | Mee eens | 58.33% | Helemaal mee eens | 16.67% | Neutraal | 16.67% | Mee oneens | 8.33% | De meerderheid is het eens met deze stelling.<br><b>Analyse:</b> In de open vragen is geen toelichting te vinden op de reactie die het niet eens was met de stelling.<br><b>Conclusie:</b> Geen aanpassing nodig. |
| Reactie  | Percentage   |  |            |            |          |        |                   |        |          |        |            |       |   |
| Mee eens   | 58.33%   |  |            |            |          |        |                   |        |          |        |            |       |   |
| Helemaal mee eens  | 16.67%   |  |            |            |          |        |                   |        |          |        |            |       |   |
| Neutraal   | 16.67%   |  |            |            |          |        |                   |        |          |        |            |       |   |
| Mee oneens   | 8.33%  |  |            |            |          |        |                   |        |          |        |            |       |   |
| <b>Kwaliteitscontroles zijn expliciet zichtbaar.</b><br><b>Weging 12,4%</b>  | 16. De methode laat zien waar in het proces kwaliteitscontroles opgenomen zijn.  | <table border="1"><thead><tr><th>Reactie</th><th>Percentage</th></tr></thead><tbody><tr><td>Mee eens</td><td>58.33%</td></tr><tr><td>Helemaal mee eens</td><td>16.67%</td></tr><tr><td>Neutraal</td><td>25.00%</td></tr><tr><td>Mee oneens</td><td>8.33%</td></tr></tbody></table> | Reactie    | Percentage | Mee eens | 58.33% | Helemaal mee eens | 16.67% | Neutraal | 25.00% | Mee oneens | 8.33% | De meerderheid is het eens met deze stelling.<br><b>Analyse:</b> In de open vragen is geen toelichting te vinden op de neutrale reacties.<br><b>Conclusie:</b> Geen aanpassing nodig.                             |
| Reactie  | Percentage   |  |            |            |          |        |                   |        |          |        |            |       |   |
| Mee eens   | 58.33%   |  |            |            |          |        |                   |        |          |        |            |       |   |
| Helemaal mee eens  | 16.67%   |  |            |            |          |        |                   |        |          |        |            |       |   |
| Neutraal   | 25.00%   |  |            |            |          |        |                   |        |          |        |            |       |   |
| Mee oneens   | 8.33%  |  |            |            |          |        |                   |        |          |        |            |       |   |
| <b>Het model moet ondersteunend zijn in het implementeren van kwaliteitsmaatregelen.</b><br><b>Weging 6,6%</b>   | 17. Het resultaat van de methode helpt mij in het implementeren van kwaliteitsmaatregelen zoals bijvoorbeeld het uitvoeren van een extra controle. | <table border="1"><thead><tr><th>Reactie</th><th>Percentage</th></tr></thead><tbody><tr><td>Mee eens</td><td>58.33%</td></tr><tr><td>Helemaal mee eens</td><td>25.00%</td></tr><tr><td>Neutraal</td><td>8.33%</td></tr><tr><td>Mee oneens</td><td>8.33%</td></tr></tbody></table>  | Reactie    | Percentage | Mee eens | 58.33% | Helemaal mee eens | 25.00% | Neutraal | 8.33%  | Mee oneens | 8.33% | De meerderheid is het eens met deze stelling.<br><b>Analyse:</b> In de open vragen is geen toelichting te vinden op de reactie die het niet eens was met de stelling.<br><b>Conclusie:</b> Geen aanpassing nodig. |
| Reactie  | Percentage   |  |            |            |          |        |                   |        |          |        |            |       |   |
| Mee eens   | 58.33%   |  |            |            |          |        |                   |        |          |        |            |       |   |
| Helemaal mee eens  | 25.00%   |  |            |            |          |        |                   |        |          |        |            |       |   |
| Neutraal   | 8.33%  |  |            |            |          |        |                   |        |          |        |            |       |   |
| Mee oneens   | 8.33%  |  |            |            |          |        |                   |        |          |        |            |       |   |

| Criterium en weging  | Vragen  | Resultaat  | Verwerking |            |                   |        |          |        |                   |        |            |        |   |
|--|---|--|------------|------------|-------------------|--------|----------|--------|-------------------|--------|------------|--------|---|
| <p><b>Het model moet het mogelijk maken om op verschillende abstractieniveaus de kwaliteit van het informatieproduct meetbaar te maken aan de hand van de relevante kwaliteitsdimensies.</b></p> <p><b>Weging 5,4%</b></p> | <p>18. Het resultaat van de methode laat mij zien wat de kwaliteit van het eindproduct is aan de hand van de relevante kwaliteitsdimensies.</p> |  <table><thead><tr><th>Reactie</th><th>Percentage</th></tr></thead><tbody><tr><td>Neutraal</td><td>50.00%</td></tr><tr><td>Mee eens</td><td>33.33%</td></tr><tr><td>Helemaal mee eens</td><td>8.33%</td></tr><tr><td>Mee oneens</td><td>8.33%</td></tr></tbody></table>    | Reactie    | Percentage | Neutraal          | 50.00% | Mee eens | 33.33% | Helemaal mee eens | 8.33%  | Mee oneens | 8.33%  | <p>Een groot deel geeft neutraal aan bij deze vraag, een minderheid is het eens is met deze stelling maar niemand is het niet eens met de stelling.</p> <p><b>Analyse:</b> In de open reacties wordt verder geen toelichting gegeven. Het criterium scoort niet slecht, maar is ook niet overtuigend. Tijdens de klankbordbijeenkomsten kwam wel naar voren dat het proces uit het onderzoek niet helemaal representatief was in de zin dat er geen “echt” product gemaakt werd, maar meer een halffabricaat dat door andere processen afgenomen wordt en verder bewerkt. Wellicht speelt dat een rol bij de beantwoording van deze vraag.</p> <p><b>Conclusie:</b> ondanks het lage gewicht van het criterium en de neutrale tot positieve score, toch opnemen als punt voor verder onderzoek. Het raakt wel de essentie van de methode. Zie ook de analyse bij het volgende punt, vraag 19.</p> |
| Reactie  | Percentage  |  |            |            |                   |        |          |        |                   |        |            |        |   |
| Neutraal   | 50.00%  |  |            |            |                   |        |          |        |                   |        |            |        |   |
| Mee eens   | 33.33%  |  |            |            |                   |        |          |        |                   |        |            |        |   |
| Helemaal mee eens  | 8.33%   |  |            |            |                   |        |          |        |                   |        |            |        |   |
| Mee oneens   | 8.33%   |  |            |            |                   |        |          |        |                   |        |            |        |   |
| <p><b>De methode moet concepten bevatten om het belang van de business impact en waarde van datakwaliteit te kunnen laten zien.</b></p> <p><b>Weging 13,4%</b></p>   | <p>19. De methode laat zien wat het belang van de impact op de business is van de kwaliteit van de data.</p>                                    |  <table><thead><tr><th>Reactie</th><th>Percentage</th></tr></thead><tbody><tr><td>Helemaal mee eens</td><td>16.67%</td></tr><tr><td>Mee eens</td><td>25.00%</td></tr><tr><td>Neutraal</td><td>25.00%</td></tr><tr><td>Mee oneens</td><td>33.33%</td></tr></tbody></table> | Reactie    | Percentage | Helemaal mee eens | 16.67% | Mee eens | 25.00% | Neutraal          | 25.00% | Mee oneens | 33.33% | <p>Voor dit punt is minder dan 50% het eens of helemaal eens met de vraag. Een derde is het oneens en neutraal.</p> <p><b>Analyse:</b> gezien de score en het gewicht van het criterium heeft dit punt aandacht. Vermoedelijk komt deze score doordat de scope van het onderzoek zich beperkt heeft tot het niveau van werkproces. Producten en diensten worden voortgebracht op het niveau van bedrijfsproces of ketenproces en door naar dat niveau te kijken kan de impact op de business beter beoordeeld worden. Uit de klankbordbijeenkomsten is al meegenomen dat de relatie van het werkproces in de context van bedrijfs- en ketenproces moet worden meegenomen.</p> <p><b>Conclusie:</b> punt wordt meegenomen in de bevindingen bij stap 2 van de methode. Toetsen in de volgende iteratie of vervolgonderzoek of het met deze aanpassing beter scoort.</p>                            |
| Reactie  | Percentage  |  |            |            |                   |        |          |        |                   |        |            |        |   |
| Helemaal mee eens  | 16.67%  |  |            |            |                   |        |          |        |                   |        |            |        |   |
| Mee eens   | 25.00%  |  |            |            |                   |        |          |        |                   |        |            |        |   |
| Neutraal   | 25.00%  |  |            |            |                   |        |          |        |                   |        |            |        |   |
| Mee oneens   | 33.33%  |  |            |            |                   |        |          |        |                   |        |            |        |   |

| Criterium en weging   | Vragen   | Resultaat   | Verwerking |            |          |        |            |        |          |        |  |
|---|--|---|------------|------------|----------|--------|------------|--------|----------|--------|--|
| <p><b>De methode moet rekening houden met de relevante context van data zodat een gedifferentieerde representatie van data mogelijk is.</b></p> <p><b>Weging 3,4%</b></p> | <p>20. De methode houdt rekening met de relevante context van data die gebruikt worden in het model.</p> | <table border="1"><thead><tr><th>Reactie</th><th>Percentage</th></tr></thead><tbody><tr><td>Mee eens</td><td>41.67%</td></tr><tr><td>Mee oneens</td><td>25.00%</td></tr><tr><td>Neutraal</td><td>33.33%</td></tr></tbody></table> | Reactie    | Percentage | Mee eens | 41.67% | Mee oneens | 25.00% | Neutraal | 33.33% | <p>Een meerderheid antwoordt positief of neutraal op deze vraag. Een kwart is het oneens met de stelling.</p> <p><b>Analyse:</b> Zie ook de vermoedelijke oorzaak zoals genoemd in vraag 14. Het procesmodel is in lagen op gebouwd. Het model dat in het onderzoek gebruikt is geeft het proces weer op processtapniveau. Op dat niveau is de gegevensbewerking niet zichtbaar. Dat is pas zichtbaar op het niveau van activiteit. Dit niveau is met de deelnemers van de klankbordgroep niet besproken.</p> <p><b>Conclusie:</b> het criterium scoort laag op belangrijkheid. De vermoedelijke verklaring vraagt niet direct om de methode aan te passen. Als in de volgende iteratie activiteitsniveau wordt meegenomen dan de vraag opnieuw evalueren.</p> |
| Reactie   | Percentage   |   |            |            |          |        |            |        |          |        |  |
| Mee eens  | 41.67%   |   |            |            |          |        |            |        |          |        |  |
| Mee oneens  | 25.00%   |   |            |            |          |        |            |        |          |        |  |
| Neutraal  | 33.33%   |   |            |            |          |        |            |        |          |        |  |